

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:16:18 ; Search time 49.45 Seconds
(without alignments)
40.806 Million cell updates/sec

Title: US-09-461-684-1
Perfect score: 109
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	100	91.7	215	2 T52523	nucleoporin p62 ho
2	100	91.7	380	2 T46395	hypothetical prote
3	100	91.7	517	2 T49173	hypothetical prote
4	93	85.3	166	2 T18513	hypothetical prote
5	90	82.6	483	2 F71619	hypothetical prote
6	87	79.8	107	2 C86477	protein F1504.29 l
7	84	77.1	441	2 A48455	acidic phosphoprot
8	81	74.3	4550	2 T18440	hypothetical prote
9	80	73.4	784	2 T18452	hypothetical prote
10	78	71.6	529	2 T50609	hypothetical prote
11	78	71.6	560	2 T06377	SAR DNA-binding pr
12	78	71.6	1560	2 T42727	proliferation pote
13	77	70.6	229	2 T07219	nuclear protein SR
14	77	70.6	266	2 A66288	protein F9L1.30 l
15	77	70.6	3724	2 T18427	hypothetical prote
16	76	69.7	511	2 S58322	nucleolar protein
17	76	69.7	812	2 S43604	R0755.1 protein (c
18	75	68.8	474	2 T38485	centromere/microtu
19	75	68.8	504	2 S48550	hypothetical prote
20	74	67.9	191	2 T23594	hypothetical prote
21	74	67.9	198	2 D97301	probable membrane
22	74	67.9	414	2 F64386	hypothetical prote
23	74	67.9	483	2 S41853	centromere/microtu
24	73	67.0	212	2 T49559	related to prolins
25	72.5	66.5	683	2 T00872	hypothetical prote
26	72	66.1	404	2 A75192	hypothetical prote
27	72	66.1	560	2 T29586	glutamic acid-rich
28	72	66.1	678	2 A54514	probable erythrocy
29	72	66.1	1701	2 T09127	

30	72	66.1	4981	2 T18489	hypothetical prote
31	71	65.1	683	2 T34103	hypothetical prote
32	71	65.1	686	2 A44842	CGMP-gated ion cha
33	71	65.1	690	2 A42161	CGMP-gated cation
34	70.5	64.7	163	2 T42696	hypothetical prote
35	70	64.2	222	2 T28919	hypothetical prote
36	70	64.2	233	2 S55165	hypothetical prote
37	70	64.2	231	2 T48617	hypothetical prote
38	70	64.2	451	2 T16418	hypothetical prote
39	70	64.2	497	2 T29814	hypothetical prote
40	70	64.2	508	2 E71620	hypothetical prote
41	70	64.2	688	2 B42161	CGMP-gated cation
42	70	64.2	2013	2 C71610	probable membrane
43	69.5	63.8	2231	2 S53416	SEN1 protein - yea
44	69	63.3	142	2 S54481	hypothetical prote
45	69	63.3	167	2 S38112	hypothetical prote

ALIGNMENTS

RESULT 1
I52523
nucleoporin p62 homolog - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C:Accession: I52523
R:Wang, Z.O.; Akmal, K.M.; Kim, K.H.
Biol. Reprod. 51, 1022-1030, 1994.
A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the
A:Reference number: I52523; MIMD:95151924
A:Accession: I52523
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-215 <RES>
A:Cross-references: GB:S75997; NID:913245; PIDN:AAB3384.1; PID:913246
A:Experimental source: testis

Query Match 91.7%; Score 100; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21
DB 35 KKKKKKKKKKKKKKKKK 54

RESULT 2
T46395
hypothetical protein DKFZp434I1120.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46395
R:Ottewill, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46395
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-380 <AAA>
A:Cross-references: EMBL:AL137556
A:Experimental source: adult testis; clone DKFZp434I1120
C:Genetics:
A>Note: DKFZp434I1120.1

Query Match 91.7%; Score 100; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KKKKKKKKKKKKKKKKK 21
IIIIIIIIIIIIIIIIIIIIII

```
Db 355 KKKKKKKKKKKKKKKKKKK 374
RESULT 3
T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C:Accession: T49173
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225017
A:Accession: T49173
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-517 <DAN>
A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A:Experimental source: cultivar Columbia; BAC clone T20N10
C:Genetics:
A:Gene: ATSP:T20N10.250
A:Map position: 3
A:Introns: 312/3: 359/3: 444/3
C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 91.7%; Score 100; DB 2; Length 517;
Best Local Similarity 100.0%; Pred. No. 0.003;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
Db 444 KKKKKKKKKKKKKKKKKKK 463

RESULT 4
T18513
hypothetical protein C0845C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T18513
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: 218935
A:Accession: T18513
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-166 <LAN>
A:Cross-references: EMBL:Z98551; PIDN:CAB1123.2
C:Genetics:
A:Map position: 3
A:Introns: 19/1
A:Note: C0845C

Query Match 85.3%; Score 93; DB 2; Length 166;
Best Local Similarity 90.0%; Pred. No. 0.0067;
Matches 18; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
Db 37 KKKKKKKKKKKKKKKKKKK 56

RESULT 5
F71619
hypothetical protein PF80235w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: F71619
R:Gardner, M.J.; Tetelji, H.; Carucci, D.J.; Cummings, L.M.; Araavind, L.; Koonin, E.V.;
Science 282: 1126-1137, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
```

```

A:Reference number: A71600; MUID:99021743
A:Accession: F71619
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-483 <GAR>
A:Cross-references: GB:AE001382; GB:AE001362; NID:93845130; PIDN:AACT1836.1; PID:938,
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF80235w

Query Match 82.6%; Score 90; DB 2; Length 483;
Best Local Similarity 85.0%; Pred. No. 0.026;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
Db 449 KKKKKKKKKKKKKKKKKKK 468

RESULT 6
C86477
protein F1504.29 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86477
R:Ritoeogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzic
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talli
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C86477
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <SNO>
A:Cross-references: GB:AE005172; NID:98778346; PIDN:AAF9354.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1504.29
A:Map position: 1

Query Match 79.8%; Score 87; DB 2; Length 107;
Best Local Similarity 94.4%; Pred. No. 0.019;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 19
Db 29 KKKKKKKKKKKKKKKKKKK 46

RESULT 7
A48455
acidic phosphoprotein PCMA1g - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: A48455
R:Deleersnijder, W.; Prasomsitt, P.; Tunpradubkul, S.; Hendrix, D.; Hamers-Casterma
Mol. Biochem. Parasitol. 56, 59-68, 1992
A:Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated
A:Reference number: A48455; MUID:93116806
A:Accession: A48455
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-441 <DEL>
A:Cross-references: GB:M95789; NID:9160602; PID:9160603
A:Experimental source: IP-PC1/C
A:Note: sequence extracted from NCBI backbone (NCBIN:121415, NCBI:P:121416)
```

C:Keywords: phosphoprotein

Query Match 77.1%; Score 84; DB 2; Length 441;
Best Local Similarity 80.0%; Pred. No. 0.089;
Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:
DB 394 KKKKKKKKKKKKKKKKE 413

RESULT 8

T18440
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18440

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18935

A:Accession: T18440

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4550 <LAW>

A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB1121.1

C:Genetics:

A:Map position: 3

A>Note: C0425w

Query Match 74.3%; Score 81; DB 2; Length 4550;
Best Local Similarity 44.2%; Pred. No. 0.77;
Matches 19; Conservative 2; Mismatches 0; Indels 22; Gaps 1;

OY 1 CRRK-----KKKKKKKKKKKKKK 21
|||||:|||||:|||||:|||||:
DB 132 CRRKNTFYNIKRYENEKYYQINNIKRRKKKKKKKKKKKK 174

RESULT 9

T18452
hypothetical protein C0560c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18452

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z18937

A:Accession: T18452

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-784 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1325245; PIDN:CAA15594.1

C:Genetics:

A:Map position: 3

A>Note: C0560c

Query Match 73.4%; Score 80; DB 2; Length 784;
Best Local Similarity 75.0%; Pred. No. 0.31;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:|||||:
DB 360 EKKKKKKKKKKKKKKKKKK 379

RESULT 10

T50609
hypothetical protein DKFP761B2423.1 - human (fragment)

C:Species: Homo sapiens (man)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: T50609
R:Bioecker, R.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, June 2000

A:Reference number: Z25143

A:Accession: T50609

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-529 <AAA>

A:Cross-references: EMBL:AJ359564

A:Experimental source: adult amygdala; clone DKFP761B2423

C:Genetics:

A>Note: DKFP761B2423.1

Query Match 71.6%; Score 78; DB 2; Length 529;
Best Local Similarity 80.0%; Pred. No. 0.37;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:|||||:
DB 464 KKKKKKKKKKKKKKKKKKK 483

RESULT 11

T06377
SAR DNA-binding protein-1 - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 02-Jun-2000
C:Accession: T06377

R:Halton, D.; Gray, J.C.

submitted to the EMBL Data Library, April 1998

A:Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nuc

A:Reference number: Z15637

A:Accession: T06377

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-560 <HAT>

A:Cross-references: EMBL:AF061962; NID:g3132695; PIDN:AA016330.1; PID:g3132696

C:Genetics:

A:Gene: SARBP-1

C:Superfamily: garden pea SAR DNA-binding protein

Query Match 71.6%; Score 78; DB 2; Length 560;
Best Local Similarity 75.0%; Pred. No. 0.38;
Matches 15; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||:|||||:
DB 463 KKKKKKKKKKKKKKKKKKK 482

RESULT 12

T42727
proliferation potential-related protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
C:Accession: T42727

R:Mitte, M.M.; Scott, R.E.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z22246

A:Accession: T42727

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1560 <RT>

A:Cross-references: EMBL:U083913; NID:g3858884; PID:g3858885; PIDN:AA072432.1

A:Experimental source: strain Balb/C

C:Genetics:

A:Gene: P2P-R

C:Function: involved in hnRNP association and Rb1 binding

C:Superfamily: RING finger homology <RRN>
F:57-107/Domain: RING finger homology <RRN>

Query Match 71.6%; Score 78; DB 2; Length 1560;
Best Local Similarity 80.0%; Pred. No. 0.74;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||
Db 1497 KKKKKKKKKKKKKKKKK 1516

RESULT 13

JC7219
nuclear protein SR-25 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: JC7219
R:Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwanana, H.; Yoshimoto, K.; Miyag
Biochem. Biophys. Res. Commun. 269, 444-450, 2000
A:Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.
A:Reference number: JC7219; MID:20175222
A:Accession: JC7219
A:Molecule type: mRNA
A:Residues: 1-229 <SAS>
A:Cross-references: DDBJ:AB035383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896
C:Experimental source: MINE cell line
C:Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine r
A:Splicing factors.
C:Keywords: nucleus; RNA processing

Query Match 70.6%; Score 77; DB 2; Length 229;
Best Local Similarity 75.0%; Pred. No. 0.27;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||
Db 95 KKKKKKKKKKKKKKKKKKK 114

RESULT 14

A86288
protein F9L1.30 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86288
R:Thoclogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;
ansen, N.E.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MID:21016719
A:Accession: A86288
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <STOC>
A:Cross-references: GB:AE005172; NID:g5103832; PIDN:AAD39662.1; GSPDB:GNO0141
C:Genetics:
A:Gene: F9L1.30
A:Map position: 1

Query Match 70.6%; Score 77; DB 2; Length 266;
Best Local Similarity 70.0%; Pred. No. 0.3;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
|||||:|||||:|||||

Db 231 KKKKKKKKKKKKKKKKK 250

RESULT 15

hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: 218935
A:Accession: T18427
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:298547; NID:e1325376; PID:e1325379; PIDN:CA81104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A>Note: C0335c

Query Match 70.6%; Score 77; DB 2; Length 3724;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CKKKKKKKKKKKKKKKKKK 20
|||||:|||||:|||||
Db 2201 CKKKKKKKKKKKKKKKK 2220

Search completed: July 1, 2002, 06:28:08
Job time: 710 sec

Mon Jul 1 13:54:45 2002

us-09-461-684-1.rpt

Page 5

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:20:23 ; Search time 25.37 Seconds

(without alignments)
32.050 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109

Sequence: 1 CKKKKKKKKKKKKKKKKKKK

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	77.1	441	1 PHPA_PLACH	002752 Plasmodium
2	76	69.7	511	1 NOP5_YEAST	012499 saccharomyc
3	75	68.8	474	1 CBP5_SCHPO	014007 schizosacch
4	75	68.8	504	1 SIK1_YEAST	012460 saccharomyc
5	75	68.8	2404	1 SON_MOUSE	090347 mus musculu
6	75	68.8	2426	1 SON_HUMAN	118583 homo sapien
7	74	67.9	414	1 Y694_METJA	058105 methanococc
8	74	67.9	474	1 CBP5_KIOLA	013473 Kluveromyc
9	74	67.9	483	1 CBP5_YEAST	033322 saccharomyc
10	74	67.9	726	1 BRD3_HUMAN	015059 homo sapien
11	72	66.1	479	1 CBP5_CANAL	043101 candida alb
12	72	66.1	678	1 GARP_PLAFA	113816 plasmodium
13	71	65.1	351	1 CG79_HUMAN	093388 homo sapien
14	71	65.1	686	1 CNG1_HUMAN	129973 h gmp-gate
15	70.5	64.7	1411	1 TCOR_HUMAN	013428 homo sapien
16	70	64.2	233	1 YJ08_YEAST	094905 saccharomyc
17	70	64.2	599	1 HM21_HUMAN	094905 homo sapien
18	70	64.2	683	1 CNG1_RAT	062927 r gmp-gate
19	70	64.2	684	1 CNG1_MOUSE	129974 m gmp-gate
20	69.5	63.8	534	1 NOP5_RAT	099286 rattus norv
21	69.5	63.8	2231	1 SEN1_YEAST	000416 saccharomyc
22	69	63.3	142	1 YMBH_YEAST	003525 saccharomyc
23	69	63.3	167	1 YK20_YEAST	003525 saccharomyc
24	69	63.3	723	1 SSRP_DROME	005344 drosophila
25	69	63.3	843	1 BLVR_BOVIN	003368 bos taurus
26	68.5	62.8	724	1 YOS1_CAEEL	034600 caenorhabd
27	68	62.4	523	1 DBP3_YEAST	120447 saccharomyc
28	68	62.4	1178	1 MNNA_YEAST	036044 saccharomyc
29	67	61.5	118	1 YQ93_CAEEL	010032 caenorhabd
30	67	61.5	690	1 CNG1_BOVIN	000194 b gmp-gate
31	67	61.5	691	1 CNG1_CANEA	028279 c gmp-gate
32	67	61.5	1002	1 IF2P_YEAST	139730 saccharomyc
33	67	61.5	1220	1 IF2P_HUMAN	060841 homo sapien

34	67	61.5	1362	1 BRD4_HUMAN	060885 homo sapien
35	66	60.6	481	1 CBP5_EMBU	043100 emeticella
36	66	60.6	487	1 CBP5_ASFPU	043102 aspergillus
37	66	60.6	687	1 YEO1_SCHPO	013702 schizosacch
38	66	60.6	1153	1 A3D1_HUMAN	014617 homo sapien
39	66	60.6	1240	1 YN11_YEAST	033935 saccharomyc
40	65.5	60.1	508	1 NO60_DROME	044081 drosophila
41	65.5	60.1	514	1 DKC1_HUMAN	060883 homo sapien
42	65	59.6	217	1 KSL_HYDAT	038978 hydra atten
43	65	59.6	271	1 YCSM_YEAST	033335 saccharomyc
44	65	59.6	320	1 YJ33_YEAST	012117 saccharomyc
45	65	59.6	344	1 YR02_YEAST	038079 saccharomyc

ALIGNMENTS

RESULT	ID	PHPA_PLACH	STANDARD:	PRT:	441 AA.
AC	002752;				
DT	01-JUL-1993 (Rel. 26, Created)				
DT	01-JUL-1993 (Rel. 26, Last sequence update)				
DT	01-JUN-1994 (Rel. 29, Last annotation update)				
DE	Acidic phosphoprotein precursor (50 kDa antigen).				
GN	PCENAL1.				
OS	Plasmodium chabaudi.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5825;				
RP	[1]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN-IP-PC1.				
RA	MEDLINE=93116806; PubMed=1475002;				
RA	Deleersnijder W., Pratsmilti P., Tungratubkul S., Hendrix D.,				
RA	Hamers-Casterman C., Hamers R.;				
RT	"Structure of a Plasmodium chabaudi acidic phosphoprotein that is				
RT	associated with the host erythrocyte membrane.";				
RL	Mol. Biochem. Parasitol. 56:59-68(1992).				
CC	- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES				
CC	THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE				
CC	PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.				
CC	- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE				
CC	CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.				
CC	- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE				
CC	THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; M95789; AAA29732.1; -				
DR	PIR; A48458; A48455.				
KW	Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.				
FT	SIGNAL	1	441	OR 24 (POTENTIAL).	
FT	CHAIN	16	441	ACIDIC PHOSPHOPROTEIN.	
FT	DOMAIN	186	313	16 X 8 AA TANDEN REPEATS.	
FT	REPEAT	186	193	1-1.	
FT	REPEAT	193	201	1-2.	
FT	REPEAT	201	209	1-3.	
FT	REPEAT	209	217	1-4.	
FT	REPEAT	217	225	1-5.	
FT	REPEAT	225	233	1-6.	
FT	REPEAT	233	241	1-7.	
FT	REPEAT	241	249	1-8.	
FT	REPEAT	249	257	1-9.	
FT	REPEAT	257	265	1-10.	
FT	REPEAT	265	273	1-11.	
FT	REPEAT	273	281	1-12.	
FT	REPEAT	281	289	1-13.	

FT REPEAT 290 297 1-14.
 FT REPEAT 298 305 1-15.
 FT REPEAT 306 313 1-16.
 FT DOMAIN 353 370 2 x 9 AA TANDEM REPEATS.
 FT REPEAT 353 360 2-1.
 FT REPEAT 361 368 2-2.
 FT DOMAIN 371 417 LYS-RICH (BASIC).
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 441 AA; 49708 MW; DB85E83E795E7E5 CRC64;

Query Match 77.1%; Score 84; DB 1; Length 441;
 Best Local Similarity 80.0%; Pred No. 0.023;
 Matches 16; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 394 KKKKKKKKKKKKKKKKK 413

RESULT 2
 NOPS_YEAST STANDARD; PRT; 511 AA.
 AC Q12499;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nucleolar protein NOP58 (Nucleolar protein NOP5).
 GN NOP58 OR NOP5 OR YOR310C OR O6108.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9288C; FY1679;
 RA Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=9288C;
 RC MEDLINE=98298165; PubMed=9632712;
 RA Wu P., Brockenbrough J.S., Metcalfe A.C., Chen S., Aris J.P.;
 RL pre-18S rRNA processing in yeast.
 RL J. Biol. Chem. 273:16453-16463(1998).
 CC -1- FUNCTION: REQUIRED FOR PRE-18S RNA PROCESSING. MAY BIND
 MICROTUBULES.
 CC -1- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- SIMILARITY: BELONGS TO THE NOPS/NOP56 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X90565; CAA62165.1; -
 CC EMBL; 275217; CAA99630.1; -
 CC EMBL; AF056070; AAC39484.1; -
 CC SCD: S0005837; NOP58.
 CC InterPro: IPR002687; NOP.
 CC Pfam: PF01796; NOP; 1.
 CC RiboDom: PD004104; NOP; 1.
 CC Ribosome biogenesis; Nuclear protein; rRNA processing.
 FT DOMAIN 441 511 ASP/GLU/LYS-RICH.
 SQ SEQUENCE 511 AA; 56956 MW; 8A2889448B2A13E2 CRC64;

Query Match 69.7%; Score 76; DB 1; Length 511;
 Best Local Similarity 70.0%; Pred. No. 0.15;
 Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 480 KKKKKKKKKKKKKKKKK 499

RESULT 3
 CBFS_SCHPO STANDARD; PRT; 474 AA.
 ID CBFS_SCHPO
 AC Q14007;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Centromere/microtubule binding protein cbf5 (centromere-binding factor
 DE 5) (Nucleolar protein cbf5).
 GN CBFS OR SPAC29A4.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churcher C.M., Wood V., Bartell B.G., Rajandream M.A.;
 RL Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBF5-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 297210; CAB10131.1; -
 CC InterPro: IPR002478; PUA.
 CC InterPro: IPR002501; TRUB_N.
 CC Pfam: PF01472; PUA; 1.
 CC Pfam: PF01509; TRUB_N; 1.
 CC SMART: SM00359; PUA; 1.
 CC Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
 FT DOMAIN 434 468 7 x 3 AA APPROXIMATE TANDEM REPEATS OF
 K-K-E.
 FT REPEAT 443 445 1.
 FT REPEAT 450 452 2.
 FT REPEAT 454 456 3.
 FT REPEAT 457 459 4.
 FT REPEAT 460 462 5.
 FT REPEAT 463 465 6.
 FT REPEAT 466 468 7.
 SQ SEQUENCE 474 AA; 53110 MW; B8C9896C5FAEB71 CRC64;

Query Match 68.8%; Score 75; DB 1; Length 474;
 Best Local Similarity 73.7%; Pred. No. 0.18;
 Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 2 KKKKKKKKKKKKKKKKKKK 20
 DB 454 KKKKKKKKKKKKKKKKK 472

RESULT 4

ID	SIKL_YEAST	STANDARD	PRI	504 AA.
AC	012460.			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	SIK1 protein (Nucleolar protein NOP56).			
GN	SIK1 OR NOP56 OR YLR197W OR I8167.9.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
CC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
CC	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YPH1;			
RX	MEDLINE=96040178; PubMed=7547500;			
RT	Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;			
RT	"Genetic analysis of growth inhibition by GAL4-L kappa B-alpha in			
RL	Saccharomyces cerevisiae."			
RL	Cell Growth Differ. 6:789-798(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Favellon A., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,			
RA	Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,			
RA	Johnson L., Langston Y., Latreille P., Mardis E., Meneses S.,			
RA	Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,			
RA	Talch S., Trevasakis E., Vaudin M., Vignati D., Wilcox L., Wilson R.,			
RA	Woldman P., Waterson R.;			
RL	Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	CHARACTERIZATION AND MUTAGENESIS.			
RX	MEDLINE=96036777; PubMed=9372940.			
RA	Gautier T., Berges T., Tollervey D., Hurt E.;			
RT	"Nucleolar KEE/D repeat proteins Nop56p and Nop58p interact with Nop1p			
RT	and are required for ribosome biogenesis."			
RL	Mol. Cell. Biol. 17:7088-7098(1997).			
CC	-1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.			
CC	-1- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.			
CC	-1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U0237; AAC4906.1; -;			
DR	EMBL; U14913; AAB67431.1; -;			
DR	SCD; S0004187; SIK1.			
DR	InterPro: IPR002687; NOP.			
DR	Pfam: PF01798; NOP. 1.			
DR	ProDom: PD004104; NOP. 1.			
KM	Ribosome biogenesis; Nuclear protein.			
FT	DOMAIN 443 504			
FT	MUTAGEN 333 353			
FT	TEMPERATURES: WHEN ASSOCIATED WITH R-385.			
FT	Y->: AT 37 DEGREES, GROWTH SLOWS AFTER 6			
FT	TO 8 HOURS AND CELL DIVISION STOPS AFTER			
FT	20 HOURS.			
FT	M->: REDUCED GROWTH RATE AT ALL			
FT	TEMPERATURES; WHEN ASSOCIATED WITH A-333.			
FT	SEQUENCE 504 AA; 56864 MM; RS522A5870DE4847 CCG64;			

Query Match	68.8%	Score 75	DB 1	Length 504
Best Local Similarity	70.0%	Pred. No. 0.19		
Matches 14	Conservative 5	Mismatches 1	Indels 0	Gaps 0

0Y 2 KKKKKKKKKKKKKKKKKKK 21
 ||||| ||||| ||||| |||||
 Db 465 KKEKKKKKKKKKKKKKK 484
 RESULT 5
 SON_MOUSE STANDARD: PRT: 2404 AA.
 ID SON_MOUSE
 AC Q90X47; Q90XP5; Q9CQK6; Q9CQ12;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE SON protein.
 GN SON.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 FN [1]
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RP STRAIN=129/Sv;
 RC MEDLINE=20400866; PubMed=10950926;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kaotaka K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guelincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL
 CC -1 FUNCTION: Transcriptional repressor. Binds to the consensus DNA
 CC sequence: 5'-GAG[GT]AN[CG][AG]CC-3'. Might protect cells from
 CC apoptosis. Might be involved in pre-mRNA splicing (By similarity).
 CC -1 SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1 ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
 CC produced by alternative splicing.
 CC -1 TISSUE SPECIFICITY: Widely expressed.
 CC -1 DOMAIN: Contains 8 types of repeats which are distributed in 3
 CC regions.
 CC -1 SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.
 CC -1 SIMILARITY: CONTAINS 1 DBRM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/announce/
 CC or_send_an_email_to_license@sib-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch)).
 CC
 CC EMBL; AF193606; AAF23120.1; -
 CC EMBL; AF193595; AAF23120.1; JOINED

DR EMBL; AF193500; AAF23120.1; -
DR EMBL; AF193595; AAF23120.1; JOINED.

DR	EMBL	AF1935596	AAF23120.1	JOINED	13 X 10 AA TANDEM REPEATS OF L-A-[ST]-[NSG]-[TS]-MDSOM.
DR	EMBL	AF1935597	AAF23120.1	JOINED	11 X 7 AA TANDEM REPEATS OF [DR]-P-Y-R-[LI][AG][OHP].
DR	EMBL	AF1935598	AAF23120.1	JOINED	14 X 6 AA REPEATS OF [ED]-R-S-M-W-S.
DR	EMBL	AF1935599	AAF23120.1	JOINED	1-1.
DR	EMBL	AF193600	AAF23120.1	JOINED	1-2.
DR	EMBL	AF193601	AAF23120.1	JOINED	1-3.
DR	EMBL	AF193602	AAF23120.1	JOINED	1-4.
DR	EMBL	AF193603	AAF23120.1	JOINED	1-5.
DR	EMBL	AF193604	AAF23120.1	JOINED	1-6.
DR	EMBL	AF193605	AAF23120.1	JOINED	1-7.
DR	EMBL	AF193607	AAF23121.1	-	1-8.
DR	EMBL	AK019312	BAB31658.1	-	1-9.
DR	EMBL	AK015081	BAB31536.1	-	1-10.
DR	EMBL	AK008478	BAB25691.1	-	1-11.
DR	EMBL	AK008256	BAB25562.1	-	1-12.
DR	MGI	MGI:98353	Son.	-	1-13.
DR	InterPro	IPR000159	DS_RBD.	-	1-14.
DR	InterPro	IPR000467	G_patch.	-	3 X 11 AA TANDEM REPEATS OF P-P-L-P-P-E-E-P-P-[TWE]-[MTC].
DR	Pfam	PF000035	dsmr.1	-	7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
DR	Pfam	PF01585	G_patch.1	-	2-1.
DR	SMART	SM00443	G_patch.1	-	2-2.
DR	PROSITE	PS00137	DS_RBD.1	-	2-3.
DR	PROSITE	PS00174	G_PATCH.1	-	2-4.
KW	RNA-binding	DNA-binding	Nuclear protein	Repeat	2-5.
KW	Alternative splicing				2-6.
FT	DOMAIN	721	850		2-7 (APPROXIMATE).
FT	DOMAIN	867	943		2 X 19 AA REPEATS OF P-S-R-R-R-S-R-S-V-V-R-R-R-S-F-S-I-S.
FT	DOMAIN	961	1080		3-1.
FT	REPEAT	961	966		3-2 (APPROXIMATE).
FT	REPEAT	969	974		3 X TANDEM REPEATS OF [ST]-P-[VLI]-R-[RL]-[RK]-[RF]-S-R.
FT	REPEAT	976	981		G-PATCH.
FT	REPEAT	985	990		DRM.
FT	REPEAT	993	998		K-> F (IN ISOFORM 2).
FT	REPEAT	1001	1006		MISSING (IN ISOFORM 2).
FT	REPEAT	1010	1015		648BF2ED3FC01D9 CRC64:
FT	REPEAT	1018	1023		
FT	REPEAT	1026	1031		
FT	REPEAT	1035	1040		
FT	REPEAT	1044	1049		
FT	REPEAT	1055	1060		
FT	REPEAT	1066	1071		
FT	REPEAT	1075	1080		
FT	DOMAIN	1101	1133		
FT	DOMAIN	1910	1979		
FT	REPEAT	1910	1916		
FT	REPEAT	1938	1944		
FT	REPEAT	1945	1951		
FT	REPEAT	1952	1958		
FT	REPEAT	1959	1965		
FT	REPEAT	1966	1972		
FT	REPEAT	1973	1979		
FT	DOMAIN	1919	1990		
FT	REPEAT	1919	1937		
FT	REPEAT	1980	1990		
FT	DOMAIN	1991	2017		
FT	DOMAIN	2283	2329		
FT	DOMAIN	2349	2404		
FT	VARSPIC	2086	2086		
FT	VARSPIC	2087	2404		
FT	SEQUENCE	2404 AA	261428 MW		

```

Query Match:      68.8%;  Score 75;  DB 1;  Length 2404;
Best Local Similarity 75.0%;  Pred. No. 0.6;
Matches 15;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0
OY 2 KKKKKKKKKKKKKKKKKKKKK 21
    ||| ||| ||| ||| ||| ||| |||

```

Dj 109 KKKKKKKKKKKKKKKEREK 128

RESULT 6
SON_HUMAN ID SON_HUMAN STANDARD; PRT; 2426 AA.
AC PI8583; O95981; GUPPYO; Q14120; O14487; Q9URP9; Q9H7B1; Q9P070;
CQ C9P072;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DEFON protein (SON3) (Negative regulatory element-binding protein) (NRE-binding protein) (DBF-5) (Box antagonist selected in saccharomyces I) (BASSI) (protein C21orf50).
GN SON OR NREP OR DBP5 OR C21ORF50 OR KIAA1019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A; B; C; D; E AND F).
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedl M., Neegaard Henriksen C., Chapot F., Deutsch S., Ucla C., Rossier C., Lyle R., Guttpont M., Antonarakis S.E.;
RT "From PRDs and open reading frames to cDNA isolation: revisiting the Human Chromosome 21 Transcription Map.";
RL Genomics 78:46-54(2001).
[2]
RP SEQUENCE FROM N.A. (ISOFORM G).
RC TISSUE=Liver;
RX MEDLINE=21316479; PubMed=11306577;
RA Sun C.-T., Lo W.-Y., Wang L.-H., Lo Y.-H., Shiou S.-R., Lai C.-K., Ting L.-P.;
RT "Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON.";
RL J Biol. Chem. 276:24059-24067(2001).
[3]
RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RC TISSUE=Placenta;
RA Casadei R., Stippoli P., D'Addabbo P., Canaider S., Lenzi L., Vitale L.R., Giannone S., Carinci P., Zannoni M.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DDJ databases.
[4]
RP SEQUENCE OF 1-130 FROM N.A.
RC TISSUE=Smooth muscle;
RA Kawakami T., Noguchi S., Itch T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujitara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Oca T., Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugino S.;
RT "NEBO human cDNA sequencing project."
RL Submitted (Aug-2000) to the EMBL/Genbank/DDJ databases.
[5]
RP SEQUENCE OF 1-114 FROM N.A.
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Gan Z.O., Wang L., Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RL "Human partial CDS from cd34+ stem cells,"
Submitted (May-1999) to the EMBL/Genbank/DDJ databases.
[6]
RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RC TISSUE=Brain;
RX MEDLINE=99357452; PubMed=10470851;
RA Kitano R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RNA Res. 6:197-205(1999).
[7]
RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RX MEDLINE=92049296; PubMed=1944235;


```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0694.
GN MJ0694.
OS Methanococcus jannaschli.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; Pubmed=8688087;
RA Bait C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.M., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschli."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67516; AAB98689.1; -
DR TIGR: MJ0694; -
DR InterPro: IPR002687; NOP.
DR Pfam: PF01798; NOP; 1.
DR ProDom: PD004104; NOP; 1.
DR Hypothetical protein; Complete proteome.
FT DOMAIN 349 414 ASP/GLU/LYS-RICH.
FT SEQUENCE 414 AA; 47799 MW; A9092EFC3C82C407 CRC64;
SQ

```

Query Match 67.9%; Score 74; DB 1; Length 414;
Best Local Similarity 75.0%; Pred. No. 0.2;
Matches 15; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 377 KKKKKKKKKKKKKKKKKKK 396

```

RESULT 8
CBF5_KLULA STANDARD; PRT; 474 AA.
AC 013473;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE 5) (Nucleolar protein CBF5).
GN CBF5.
OS Kluyveromyces fragilis (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JBD100;
RC MEDLINE=98144788; Pubmed=9483794;
RA Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,
RA Hooykaas P.J.J.;

```

RT "The lysine-rich C-terminal repeats of the centromere-binding factor
RT 5 (Cbf5) of Kluyveromyces fragilis are not essential for function."
RL Yeast 14:37-48(1998).
CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A) (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TUB FAMILY OF PSEUDODUPLEX SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF008563; AAC64862.1; -
DR InterPro: IPR002478; PUB_N.
DR InterPro: IPR002501; TUB_N.
DR Pfam: PF01472; PUB; 1.
DR Pfam: PF01509; TUB_N; 1.
DR SMART: SM00359; PUB; 1.
KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 431 460
FT REPEAT 431 433 1.
FT REPEAT 434 436 2.
FT REPEAT 437 439 3.
FT REPEAT 440 442 4.
FT REPEAT 443 445 5.
FT REPEAT 446 448 6.
FT REPEAT 449 451 7.
FT REPEAT 452 454 8.
FT REPEAT 455 457 9.
FT REPEAT 458 460 10.
FT SEQUENCE 474 AA; 53630 MW; 95306E7FEA756C CRC64;
SQ

```

Query Match 67.9%; Score 74; DB 1; Length 474;
Best Local Similarity 70.0%; Pred. No. 0.22;
Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 KKKKKKKKKKKKKKKKKKKKK 21
DB 434 KKKKKKKKKKKKKKKKKKK 453

```

RESULT 9
CBF5_YEAST STANDARD; PRT; 483 AA.
AC P33322;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Centromere/microtubule binding protein CBF5 (centromere-binding factor
DE 5) (Nucleolar protein CBF5) (P64').
GN CBF5 OR YLR175W OR U9470.11.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC MEDLINE=9330283; Pubmed=8336724;
RA Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.,
RA "An essential yeast protein, CBF5p, binds in vitro to centromeres and
RA microtubules." J.
RA Mol. Cell. Biol. 13:4884-4893(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RC Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,

RA Favellio A., Fulton L., Gatling S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Riken L., Riles L., Taich A., Trevasakis E., Vignati D.,
 RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.,
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
 CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L12351; AAA34473.1; -
 DR EMBL: U17246; AAB67463.1; -
 DR PIR: S41853; S41853.
 DR SCD: S0004165; CBF3.
 DR InterPro: IPR002478; CBF3.
 DR InterPro: IPR002501; Trub_N.
 DR Pfam: PF01472; PUA; 1.
 DR Pfam: PF01509; Trub_N; 1.
 DR SMART: SM00359; PUA; 1.
 KM Microtubules: Centromere; Repeat; Nuclear protein; DNA-binding.
 FT DOMAIN 434 463 10 X 3 AA TANDEM REPEATS OF K-K-IDE)
 FT REPEAT 434 436 1.
 FT REPEAT 437 439 2.
 FT REPEAT 440 442 3.
 FT REPEAT 443 445 4.
 FT REPEAT 446 448 5.
 FT REPEAT 449 451 6.
 FT REPEAT 452 454 7.
 FT REPEAT 455 457 8.
 FT REPEAT 458 460 9.
 FT REPEAT 461 463 10.
 SQ SEQUENCE 483 AA; 54704 MW; D356B39FDCC32E2D CRC64;

Query Match 67.9%; Score 74; DB 1; Length 483;
 Best Local Similarity 70.0%; Pred. No. 0.23;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 434 KKEKKKKKKKKKKKKKK 453

RESULT 10
 BRD3_HUMAN STANDARD; PRT; 726 AA.
 AC 015059; 092645;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bromodomain-containing protein 3 (RING3-like protein).
 GN BRD3 OR RING3L OR KIAA0043.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051398; PubMed=7584044;
 RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,

RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.,
 RT "Prediction of the coding sequences of unidentified human genes. II.
 RT The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by
 RT analysis of cDNA clones from human cell line KG-1."
 RL DNA Res. 1:223-229(1994).
 RN [2]
 RP SEQUENCE OF 363-726 FROM N.A.
 RA MEDLINE=98038990; PubMed=9373153;
 RA Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowdale J., Beck S.,
 RT "Chromosomal localization, gene structure and transcription pattern of
 RT the ORFX gene, a homologue of the MHC-linked RING3 gene."
 RL Gene 200:177-183(1997).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D26362; BA405393.1; -
 DR EMBL: Z81330; CAB03630.1; -
 DR MIM: 601541; -
 DR InterPro: IPR001487; Bromodomain.
 DR Pfam: PF00439; bromodomain; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 2.
 DR PROSITE: PS00633; BROMODOMAIN_1; 2.
 DR PROSITE: PS50014; BROMODOMAIN_2; 2.
 KM Bromodomain; Repeat; Nuclear protein.
 FT DOMAIN 56 115 115
 FT DOMAIN 326 398 BROMODOMAIN 1.
 FT DOMAIN 487 555 BROMODOMAIN 2.
 FT DOMAIN 676 725 LYS-RICH.
 FT CONFLICT 465 466 SER-RICH.
 FT SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 67.9%; Score 74; DB 1; Length 726;
 Best Local Similarity 70.0%; Pred. No. 0.31;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 489 KKKKKKKKKKKKKKKKK 508

RESULT 11
 CBF5_CANAL STANDARD; PRT; 479 AA.
 AC 043101;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Centromere/microtubule binding protein CBF5 (Centromere-binding factor
 DE 5) (Nucleolar protein CBF5).
 GN Candida albicans (Yeast).
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 CC NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jiang W., Clifford J., Koltin Y.,
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
 CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
 CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
 CC SOME WAY ASSOCIATED WITH THE CBF3 110 KDA SUBUNIT (CBF3A). (BY

CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U59149; AAB94297.1; -
DR InterPro: IPR002478; PUA.
DR InterPro: IPR002501; TRUB.N.
DR Pfam: PF01472; PUA; 1.
DR Pfam: PF01509; TRUB.N; 1.
DR SMART: SM00359; PUA; 1.
KM Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
SQ SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 66.1%; Score 72; DB 1; Length 479;
Best Local Similarity 70.0%; Pred. No. 0.35;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21
DB 432 KKKKKKKKKKKKKKK 451

RESULT 12
ID GARP_PLAF STANDARD; PRT; 678 AA.
AC P13616; -
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
CX NCBI_TaxID=5637;
RN [1]
RP MEDLINE=69040048; PubMed=2903445;
RA Trigila T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP).";
RL Mol. Biochem. Parasitol. 51:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03998; AAA29605.1; -
DR PIR: A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF R-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; ZA8F85606496EA9E CRC64;

Query Match 66.1%; Score 72; DB 1; Length 678;
Best Local Similarity 70.0%; Pred. No. 0.46;
Matches 14; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 21
DB 132 KKKKKKKKKKKKKKK 151

RESULT 13
ID CG79_HUMAN STANDARD; PRT; 351 AA.
AC Q9Y388;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CG7-79.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics";
RL Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: CONTRAINS A RNA RECOGNITION MOTIF (RNM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF151837; AAD34074.1; -
DR HSP; P09651; 1HA1.
DR InterPro: IPR000504; RNM.
DR Pfam: PF00076; RNM; 1.
DR SMART: SM00360; RNM; 1.
DR PROSITE: PS50102; RNM; 1.
DR PROSITE: PS00030; RNM; RNP; 1;
KM Hypothetical protein; RNA-binding.
FT DOMAIN 36 114 RNA-BINDING (RNM).
SQ SEQUENCE 351 AA; 39675 MW; 756E882DB192EBE CRC64;

Query Match 65.1%; Score 71; DB 1; Length 351;
Best Local Similarity 73.7%; Pred. No. 0.35;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKK 20
DB 156 KKKKKKKKKKKKKKK 174

RESULT 14
ID CNG1_HUMAN STANDARD; PRT; 686 AA.
AC P29973; Q16485; Q16279;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE (Cyclic nucleotide-gated channel alpha 1) (Cyclic nucleotide-gated
DE channel, photoreceptor) (Cyclic nucleotide-gated cation channel 1)
DE (Rod photoreceptor cGMP-gated channel alpha subunit).
GN CNGA1 OR CNCG1 OR CNCG.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA MEDLINE=92210603; PubMed=1372902;
 RA Piltner S.J., Lee A.K., Altherr M.R., Howard T.A., Seidlin M.F.,
 RA Hurvitz R.L., Wasmuth J.J., Baehr W.;
 RT "Primary structure and chromosomal localization of human and mouse
 RT rod photoreceptor cGMP-gated cation channel.";
 RL J. Biol. Chem. 267:6257-6262(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA MEDLINE=92356211; PubMed=1379636;
 RA Dhallan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,
 RA Yau K.-W., Nathans J.;
 RT "Human rod photoreceptor cGMP-gated channel: amino acid sequence,
 RT gene structure, and functional expression.";
 RL J. Neurosci. 12:3248-3256(1992).
 RN [3]
 RP SEQUENCE OF 313-573 FROM N.A.
 RA MEDLINE=95175019; PubMed=7532814;
 RA Distler M., Biel M., Flockert V., Hofmann F.;
 RT "Expression of cyclic nucleotide-gated cation channels in non-sensory
 RT tissues and cells.";
 RL Neuropharmacology 33:1275-1282(1994).
 RN [4]
 RP VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.
 RA MEDLINE=96036047; PubMed=7479749;
 RA Drya T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;
 RT "Mutations in the gene encoding the alpha subunit of the rod
 RT cGMP-gated channel in autosomal recessive retinitis pigmentosa.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).
 CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
 CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
 CC PHOTORECEPTORS.
 CC -1- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
 CC COMPLEX WITH CNGB.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
 CC -1- DISEASE: DEFECTS IN CNGB1 ARE A CAUSE OF AUTOSOMAL RECESSIVE
 CC RETINITIS PIGMENTOSA (ARRP). A DISEASE THAT LEADS TO DEGENERATION
 CC OF RETINAL PHOTORECEPTOR CELLS.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 CC FAMILY.
 CC -1- DATABASE: NAME=Mutations of the CNGB1 gene;
 CC NOTE=Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/cnbgmut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as their content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: W64741; AAA52010.1; ALT_INIT.
 DR EMBL: S42457; AAB22778.1; -
 DR EMBL: S76062; AAD14206.1; -
 DR PIR: A42161; A42161.
 DR MIM: 123825; -
 DR InterPro: IPR000636; Cation_chan_non_lig.
 DR InterPro: IPR000595; CNMP_binding.
 DR Pfam: PF00027; CNMP_binding.1.
 DR Pfam: PF00520; Ion_trans.1.
 DR SMART: SM00100; CNMP.1.
 DR PROSITE: PS00888; CNMP_BINDING_1; 1.
 DR PROSITE: PS00889; CNMP_BINDING_2; 1.

DR PROSITE: PS0042; CNMP_BINDING_3; 1.
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;
 KW Multigene family; Vision; Disease mutation; Polymorphism;
 KW Retinitis pigmentosa.
 FT DOMAIN 1 160
 FT TRANSSEM 161 181
 FT DOMAIN 182 194
 FT TRANSSEM 195 213
 FT DOMAIN 214 237
 FT TRANSSEM 238 257
 FT DOMAIN 258 295
 FT TRANSSEM 296 318
 FT DOMAIN 319 370
 FT TRANSSEM 371 390
 FT DOMAIN 391 474
 FT TRANSSEM 475 495
 FT DOMAIN 496 686
 FT NP_BIND 483 605
 FT BINDING 542 542
 FT BINDING 557 557
 FT CARBOHYD 421 421
 FT VARIANT 28 28
 FT VARIANT 114 114
 FT VARIANT 316 316
 FT CONFLICT 46 46
 FT CONFLICT 85 85
 FT CONFLICT 146 147
 FT CONFLICT 539 539
 FT CONFLICT 677 678
 FT SEQUENCE 686 AA; 79126 MW; E520D216FC97AB6 CRC64;
 SQ
 Query Match 65.1%; Score 71; DB 1; Length 686;
 Best Local Similarity 70.0%; Pred. No. 0.57;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 OY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 124 KKKKKKKKKKKKKKKKK 143
 RESULT 15
 TCOF_HUMAN STANDARD; PRT; 1411 AA.
 ID TCOF_HUMAN
 AC Q13428; Q59408;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Treacher protein (Treacher collins syndrome protein).
 GN TCOF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96134183; PubMed=8563749;
 RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
 RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
 RT "Positional cloning of a gene involved in the pathogenesis of
 RT Treacher collins syndrome. The Treacher collins Syndrome
 RT Collaborative Group.";
 RL Nat. genet. 12:130-136(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97250498; PubMed=9096354;
 RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
 RA Ashley J.A., Lovett M., Jabs E.W.;
 RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
 RT mutations in Treacher Collins syndrome throughout its coding

RT region." ;
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
 RN [3]
 RP VARIANTS L-439; V-810; V-1313 AND G-1355, AND VARIANT TCS R-53.
 RX MEDLINE=97195537; PubMed=9042910;
 RA Edwards S.J., Gladwin A.J., Dixon M.J.;
 RT "The mutational spectrum in Treacher Collins syndrome reveals a
 RT predominance of mutations that create a premature-termination
 RT codon." ;
 RL Am. J. Hum. Genet. 60:515-524(1997).
 CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS
 CC SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF
 CC CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000
 CC LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY
 CC SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,
 CC ATRESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE
 CC EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)
 CC LATERAL DOWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH
 CC COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND
 CC ZYGOMATIC COMPLEX; (4) CLEFT PALATE
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; U40847; AAC50903.1; -;
 DR EMBL; U76366; AAC51181.1; -;
 DR EMBL; U84664; AAC51185.1; JOINED.
 DR EMBL; U84640; AAC51185.1; JOINED.
 DR EMBL; U84641; AAC51185.1; JOINED.
 DR EMBL; U84642; AAC51185.1; JOINED.
 DR EMBL; U84643; AAC51185.1; JOINED.
 DR EMBL; U84644; AAC51185.1; JOINED.
 DR EMBL; U84645; AAC51185.1; JOINED.
 DR EMBL; U84646; AAC51185.1; JOINED.
 DR EMBL; U84647; AAC51185.1; JOINED.
 DR EMBL; U84648; AAC51185.1; JOINED.
 DR EMBL; U84649; AAC51185.1; JOINED.
 DR EMBL; U84650; AAC51185.1; JOINED.
 DR EMBL; U84651; AAC51185.1; JOINED.
 DR EMBL; U84652; AAC51185.1; JOINED.
 DR EMBL; U84653; AAC51185.1; JOINED.
 DR EMBL; U84654; AAC51185.1; JOINED.
 DR EMBL; U84655; AAC51185.1; JOINED.
 DR EMBL; U84656; AAC51185.1; JOINED.
 DR EMBL; U84657; AAC51185.1; JOINED.
 DR EMBL; U84658; AAC51185.1; JOINED.
 DR EMBL; U84659; AAC51185.1; JOINED.
 DR EMBL; U84660; AAC51185.1; JOINED.
 DR EMBL; U84661; AAC51185.1; JOINED.
 DR EMBL; U84662; AAC51185.1; JOINED.
 DR EMBL; U84663; AAC51185.1; JOINED.
 DR MIM: 154500; -;
 DR InterPro: IPR003993; treeacle.
 DR PRINTS: PR01503; TREACLE.
 KW Disease mutation; Polymorphism.
 FT DOMAIN 89 97 POLY-GLU.
 FT DOMAIN 204 207 POLY-SER.
 FT DOMAIN 616 619 POLY-SER.
 FT DOMAIN 919 924 POLY-SER.
 FT DOMAIN 1285 1289 POLY-LYS.
 FT DOMAIN 1375 1386 POLY-LYS.
 FT DOMAIN 1398 1405 POLY-LYS.
 FT VARIANT 53 53 W -> R (IN TCS).
 FT VARIANT 439 439 /FTId=VAR_005630.
 FT P -> L.
 FT A -> V.
 FT VARIANT 810 810 /FTId=VAR_005631.
 FT /FTId=VAR_005632.

FT VARIANT 1313 1313 A -> V.
 FT FT /FTId=VAR_005633.
 FT VARIANT 1355 1355 D -> G.
 FT FT /FTId=VAR_005634.
 SQ SEQUENCE 1411 AA: 144312 MW: 3880203D985C2699 CRC64;
 Query Match 64.7% Score 70.5; DB 1; Length 1411;
 Best Local Similarity 54.8%; Pred. No. 1.1;
 Matches 17; Conservative 2; Mismatches 1; Indels 11; Gaps 1;
 QY 2 KKKKKKKKKK-----KKKKKK 21
 DB 1375 KKKKKKKKKAKASTKDSKSPSKKKKKK 1405
 Search completed: July 1, 2002, 06:31:33
 Job time: 670 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:28 ; Search time 83.98 Seconds

(without alignments)
43.239 Million cell updates/sec

Title: US-09-461-684-1
Perfect score: 109
Sequence: 1 CXXXXXXXXXXXXXXXXXXXX 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mbc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.protist:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriophage:*
- 17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	91.7	128	3	Q9P529
2	100	91.7	129	11	Q9P529
3	100	91.7	168	4	Q9H5V5
4	100	91.7	215	11	Q64075
5	100	91.7	260	10	Q9L629
6	100	91.7	380	4	Q9NT34
7	100	91.7	517	10	Q9LXR2
8	100	91.7	531	6	Q9SLV5
9	100	91.7	720	4	Q9H607
10	97	89.0	667	4	Q9H648
11	90	82.6	257	4	Q9H5Y3
12	90	82.6	483	5	Q96148
13	90	82.6	686	4	Q9NKF0
14	89	81.7	83	11	Q9ER82
15	87	79.8	107	10	Q9LOR6
16	81	74.3	4550	5	Q77336

Result No.	Score	Query Match	Length	DB ID	Description
17	80	73.4	130	4	Q9H7E1
18	80	73.4	213	11	P97762
19	80	73.4	784	5	Q77339
20	79	72.5	260	11	Q991X5
21	78	71.6	263	4	Q9BU76
22	78	71.6	407	4	Q00536
23	78	71.6	529	4	Q9NPX4
24	78	71.6	560	10	Q65334
25	78	71.6	628	4	Q9H5M5
26	78	71.6	1591	11	P97868
27	78	71.6	1616	4	Q96PH3
28	77.5	71.1	353	4	Q9HBO9
29	77	70.6	103	5	Q9N8X7
30	77	70.6	229	11	Q9JMG3
31	77	70.6	266	10	Q9X134
32	77	70.6	402	4	Q9H4Q7
33	77	70.6	577	3	Q9C1W8
34	77	70.6	749	5	Q967D9
35	77	70.6	3724	5	Q77320
36	76	69.7	581	5	Q96549
37	76	69.7	812	5	Q21827
38	76	69.7	1629	5	Q9U0K9
39	75	68.8	2086	11	Q9QXP5
40	75	68.8	2404	11	Q9QX47
41	74	67.9	191	5	Q45682
42	74	67.9	198	16	Q97E51
43	74	67.9	244	11	Q9CWL7
44	74	67.9	256	11	Q9CUI4
45	74	67.9	550	5	Q9VVF1

ALIGNMENTS

RESULT 1

Q9P529 ID Q9P529 PRELIMINARY; PRT; 128 AA.

AC Q9P529; ID Q9P529; PRT; 128 AA.

DT 01-Oct-2000 (TREMBLrel. 15, Created)

DT 01-Dec-2001 (TREMBLrel. 19, Last sequence update)

DT 01-Dec-2001 (TREMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 15.2 KDA PROTEIN.

GN B24H17.160.

GN Neurospora crassa.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes.

OC Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141.

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Alpn V., Hohnsels J., Brandt P., Fartmann B., Holland R., Nyakatura G., Mewes H.W., Mannhaupt G.,

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL356815; CAB92638.2; -

KW Hypothetical protein

SO SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 91.7%; Score 100; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 9.6e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KXXXXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXXXXX|

Db 71 KXXXXXXXXXXXXXXXXXXXX 90
|XXXXXXXXXXXXXXXXXXXX|

RESULT 2
Q35807 ID Q35807 PRELIMINARY; PRT; 129 AA.

AC 035807;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.
 GN MDC2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EPIDIDYMIS;
 RX MEDLINE=98172708; PubMed=9511718;
 RA Proels F., Loser B., Marx M.,
 RT "Differential expression of osteopontin, PC4, and CECS, a novel mRNA
 species, during in vitro angiogenesis.";
 RL Exp. Cell Res. 239:1-10(1998).
 DR EMBL; Y08769; CAA70022.1; -;
 DR InterPro; IPR000719; Euk_pkinase.
 DR Pfam; PF00069; pkinase.1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 129 AA; 15080 MW; 38102272B8E2EDB4 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 129;
 Best Local Similarity 100.0%; Pred. No. 9,7e-05;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 ID 85 KKKKKKKKKKKKKKKKKKK 104

RESULT 3
 Q9H5V6 PRELIMINARY; PRT; 168 AA.
 ID Q9H5V6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK026629; BAB15513.1; -;
 FT NON_TER 168
 SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 91.7%; Score 100; DB 4; Length 168;
 Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 140 KKKKKKKKKKKKKKKKKKK 159

RESULT 4
 Q64075 PRELIMINARY; PRT; 215 AA.
 ID Q64075;
 AC Q64075;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).
 OS Rattus sp.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NX NCBI_TaxID=10118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95151924; PubMed=7849178;
 RA Wang Z.Q., Akmal K.M., Kim K.H.;
 RT "An unusual nucleoporin-related messenger ribonucleic acid is present
 in the germ cells of rat testis.";
 RL Biol. Reprod. 51:1022-1030(1994).
 DR EMBL; S75997; AAB3384.1; -;
 KW Porin.
 FT NON_TER 1
 SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD8 CRC64;

Query Match 91.7%; Score 100; DB 11; Length 215;
 Best Local Similarity 100.0%; Pred. No. 0.00014;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 35 KKKKKKKKKKKKKKKKKKK 54

RESULT 5
 Q9LG29 PRELIMINARY; PRT; 260 AA.
 ID Q9LG29;
 AC Q9LG29;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FLD5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Nakamura Y.;
 RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002460; BA97098.1; -;
 DR InterPro; IPR001386; linker_histone.
 DR PRINTS; PRO0624; HISTONEH5.
 SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;

Query Match 91.7%; Score 100; DB 10; Length 260;
 Best Local Similarity 100.0%; Pred. No. 0.00016;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 7 KKKKKKKKKKKKKKKKKKK 26

RESULT 6
 Q9NT34 PRELIMINARY; PRT; 380 AA.
 ID Q9NT34;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).
 GN DKF2P434I1120.
 OS Homo sapiens (Human).

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ohtenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL137556; CAB70810.1; -
KM Hypothetical protein.
FT NON_TER 380
SQ SEQUENCE 380 AA; 42689 MW; 67F5DD101346AFB CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 4; Length 380;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 355 KKKKKKKKKKKKKKKKKKK 374

RESULT 7
Q9LXR2 PRELIMINARY; PRT; 517 AA.
AC Q9LXR2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE HYPOTHETICAL 59.7 KDA PROTEIN.
GN T20N10_250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Quettier F., Salanoubat M.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL353032; CAB88307.1; -
DR InterPro: IPR001810; F-box.
DR Pfam: PF00646; F-box.1.
DR SMART: SM00256; FBOX.1.
DR PROSITE: PS50181; FBOX.1.
KM Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 10; Length 517;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 444 KKKKKKKKKKKKKKKKKKK 463

RESULT 8
Q95LV6 PRELIMINARY; PRT; 531 AA.
AC Q95LV6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).
OX NCBI_TaxID=9606;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AB071085; BAB64479.1; -
KM Hypothetical protein.
FT NON_TER 531
SQ SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 6; Length 531;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 502 KKKKKKKKKKKKKKKKKKK 521

RESULT 9
Q9H607 PRELIMINARY; PRT; 720 AA.
AC Q9H607;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CNA: FLJ21979 FIS. CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AK025632; BAB15196.1; -
FT NON_TER 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match
Best Local Similarity 91.7%; Score 100; DB 4; Length 720;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
DB 692 KKKKKKKKKKKKKKKKKKK 711

RESULT 10
Q9HC48 PRELIMINARY; PRT; 667 AA.
AC Q9HC48;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CCLL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=21143360; PubMed=1149944;
 RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
 RA Schandendorf D.;
 RT "Serological detection of cutaneous T-cell lymphoma-associated
 RT antigens";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
 DR EMBL: AF177228; FAG3676.1; -.
 DR HSSP: Q12923; 3PDZ.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 FT NON_TER 1 1
 FT NON_TER 667 667
 SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;

Query Match 89.0%; Score 97; DB 4; Length 667;
 Best Local Similarity 95.0%; Pred. No. 0.00063;
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 648 RKKKKKKKKKKKKKKKKKK 667

RESULT 11
 Q9H3Y3 ID Q9H5Y3 PRELIMINARY; PRT; 257 AA.
 AC Q9H5Y3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CDNA: FLJ22789 FIS, CLONE KAI1A2171 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TISSUE=TESTIS;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Taniguchi A., Fujiwara T., Ono T.,
 RA Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
 RA Isogai T., Sugano S.;
 RT "NDO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026442; BAB15484.1; -.
 FT NON_TER 257 257
 FT NON_TER 257 257
 SQ SEQUENCE 257 AA; 28778 MW; 100DEA6ABA521868 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 257;
 Best Local Similarity 100.0%; Pred. No. 0.0019;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 19
 DB 240 KKKKKKKKKKKKKKKKKKK 257

RESULT 12
 Q96148 ID Q96148 PRELIMINARY; PRT; 483 AA.
 AC Q96148;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE HYPOHETICAL 57.8 KDA PROTEIN.
 GN PRB0235W.

OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Jettelin H., Carucci D.J., Cummings L.M., Aravind L.,
 RA Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J.,
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perera M.,
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
 RT falciparum";
 RL Science 282:1126-1132(1999).
 DR EMBL: AE001362; AAC71836.1; -.
 DR Hypothetical protein.
 KW SEQUENCE 483 AA; 57785 MW; 9332C1EC9A84DECFB CRC64;

Query Match 82.6%; Score 90; DB 5; Length 483;
 Best Local Similarity 85.0%; Pred. No. 0.0029;
 Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 449 KKKKKKKKKKKKKKKKKKK 468

RESULT 13
 Q9NXP0 ID Q9NXP0 PRELIMINARY; PRT; 686 AA.
 AC Q9NXP0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CDNA: FLJ20288 FIS, CLONE HEP04414 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000295; BAA91063.1; -.
 DR HSSP: O00420; IAWC.
 DR InterPro: IPR002110; ANK.
 DR Pfam: PF00023; ANK; 14.
 DR SMART: SM00248; ANK; 14.
 DR PROSITE: PS50088; ANK_REPEAT; 13.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 KW ANK repeat; Repeat.
 FT NON_TER 686 686
 FT NON_TER 686 686
 SQ SEQUENCE 686 AA; 73615 MW; AE1390212E0FAD22 CRC64;

Query Match 82.6%; Score 90; DB 4; Length 686;
 Best Local Similarity 85.0%; Pred. No. 0.0037;
 Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 KKKKKKKKKKKKKKKKKKK 21
 DB 666 KKKKKKKKKKKKKKKKKKK 685

RESULT 14
 Q9ER82 ID Q9ER82 PRELIMINARY; PRT; 83 AA.
 AC Q9ER82;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)


```
Search completed: July 1, 2002, 06:30:54
Job time: 686 sec
```

DT 01-MAR-2001 (TREMBLrel, 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel, 16, last annotation update)
DE SR528 PROTEIN (FRAGMENT).
GN SR528.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP
RP STRAIN=C57BL/10; TISSUE=SKELTAL MUSCLE;
RC
RA Kemp T.J., Sackusky T.J., Carey N., Coulton G.R.;
RT "Murine genes identified from skeletal muscle mRNA which had undergone
7 days of passive stretch in vivo.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RL EMBL: A0750693; CAC08905.1; -.
FR
FR NON_TER 1
SQ
SQ SEQUENCE 83 AA; 10093 MW; 308BB2113D02DF92 CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:10:53 ; Search time 98.97 Seconds
(without alignments)
23.568 Million cell updates/sec

Title: US-09-461-684-1

Perfect score: 109
Sequence: 1 CRRKKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: A.Genesec.032802:*
2: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1980.DAT.*
3: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1981.DAT.*
4: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1982.DAT.*
5: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1983.DAT.*
6: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1984.DAT.*
7: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1985.DAT.*
8: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1986.DAT.*
9: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1987.DAT.*
10: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1988.DAT.*
11: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1989.DAT.*
12: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1990.DAT.*
13: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1991.DAT.*
14: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1992.DAT.*
15: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1993.DAT.*
16: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1994.DAT.*
17: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1995.DAT.*
18: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1996.DAT.*
19: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1997.DAT.*
20: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1998.DAT.*
21: /SIDS1/gcgdata/genesec/genesecp-emb1/AA1999.DAT.*
22: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2000.DAT.*
23: /SIDS1/gcgdata/genesec/genesecp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	21	21	AAAB13780
2	109	100.0	39	19	AAAB15801
3	109	100.0	58	22	AAU18238
4	109	100.0	70	22	AAO11210
5	109	100.0	81	22	AAO03766
6	105	96.3	36	22	AAO11849
7	105	96.3	47	22	AAO12447
8	105	96.3	66	22	AAU18239
9	105	96.3	69	21	AAAB53800
10	105	96.3	83	22	AAAG73729
11	103	94.5	42	22	AAO09258

12	103	94.5	68	22	AAO11214	Human polypeptide
13	103	94.5	124	22	AAO00291	Human polypeptide
14	102	93.6	41	22	AAO12203	Human polypeptide
15	102	93.6	60	22	AAO12187	Human polypeptide
16	102	93.6	62	21	AAO42889	Human OREF ORF2653
17	102	93.6	69	22	AAAG74650	Human colon cancer
18	102	93.6	108	22	AAU18184	Novel human DNA-bi
19	102	93.6	108	22	AAAG0618	Human immune/haema
20	102	93.6	113	22	AAO00922	Human polypeptide
21	101	92.7	54	22	AAU18259	Novel human DNA-bi
22	101	92.7	59	22	AAU18259	Novel human DNA-bi
23	100	91.7	20	3	AAAP20159	Sequence of lysine
24	100	91.7	22	18	AAAB38833	Delivery peptide u
25	100	91.7	22	18	AAAB38796	Delivery peptide u
26	100	91.7	23	18	AAAB38877	Delivery peptide u
27	100	91.7	23	18	AAAB38834	Delivery peptide u
28	100	91.7	23	18	AAAB38797	Delivery peptide u
29	100	91.7	23	22	AAO05384	Human polypeptide
30	100	91.7	24	18	AAAB38878	Delivery peptide u
31	100	91.7	24	18	AAAB38835	Delivery peptide u
32	100	91.7	24	18	AAAB38798	Delivery peptide u
33	100	91.7	25	18	AAAB38879	Delivery peptide u
34	100	91.7	25	18	AAAB38836	Delivery peptide u
35	100	91.7	25	18	AAAB38799	Delivery peptide u
36	100	91.7	26	18	AAAB38880	Delivery peptide u
37	100	91.7	26	18	AAAB38837	Delivery peptide u
38	100	91.7	26	18	AAAB38800	Delivery peptide u
39	100	91.7	26	22	AAO04756	Human polypeptide
40	100	91.7	26	22	AAO08995	Human polypeptide
41	100	91.7	26	22	AAO09001	Human polypeptide
42	100	91.7	27	18	AAAB38838	Delivery peptide u
43	100	91.7	27	18	AAAB38801	Delivery peptide u
44	100	91.7	27	18	AAAB38881	Delivery peptide u
45	100	91.7	27	21	AAAB59105	Breast and ovarian

ALIGNMENTS

RESULT	1	
AAAB13780	AAAB13780 standard; peptide: 21 AA.	
ID	AAAB13780;	
AC	AAAB13780;	
XX	10-NOV-2000 (first entry)	
DT		
DE	Soluble peptide antigen pK.	
XX		
KW	pK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;	
KW	major histocompatibility complex class 1; MHC class 1; antigen; tumour;	
KW	prostate; breast; multiple myeloma.	
XX		
OS	Unidentified.	
XX		
PN	WO200035949-A1.	
PD	22-JUN-2000.	
XX		
PF	14-DEC-1999; 99WO-US29724.	
XX		
PR	14-DEC-1998; 98US-0112324.	
XX		
PA	(DEND-) DENDREON CORP.	
XX		
PI	Laus R, Hakim I, Vidovic D;	
XX		
DR	WPL; 2000-442365/38.	
XX		
PT	Antigens modified by the covalent addition of a peptide that	
PT	facilitates entry into antigen presenting cells, useful for producing	
PT	compositions for immunizing against tumors and pathogens -	

PS Claim 2; Page 26; 34pp; English.

XX The present invention relates to compositions of modified soluble protein

CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)

CC response i.e. a major histocompatibility complex (MHC) class I molecule

CC response. The protein antigen is modified by the covalent addition of a

CC peptide sequence which facilitate entry of the antigen into antigen

CC presenting cells (APCs). The present sequence is one such peptide

CC sequence which can be used to modify the soluble antigens. The present

CC sequence is peptide PK. The modified antigen composition may be used for

CC immunising against, or treating a tumour e.g. prostate and breast

CC carcinoma or multiple myeloma, or pathogen in mammals.

XX

SO Sequence 21 AA:

Query Match 100.0%; Score 109; DB 21; Length 21;

Best Local Similarity 100.0%; Pred. No. 2.3e-06;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21

DB 1 CXXXXXXXXXXXXXXXXX 21

RESULT 2

AAW45801

ID AAW45801 standard; peptide: 39 AA.

XX

AC AAW45801;

XX

DT 25-JUN-1998 (first entry)

XX

DE One chain of a bombesin dimer.

XX

KM Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;

KM alpha-MSH-ANT; bombesin; dimer; Divalent agonist; disulphide bond;

KM G-protein coupled receptor.

XX

OS Synthetic.

XX

FT Key Location/Qualifiers

FT Cross-links 1

FT /note= "This residue is disulphide bonded to the

FT corresponding Cys residue of an identical chain

FT to form a dimer"

FT Modified-site 28

FT /note= "Epsilon-aminohexanoic acid"

PN

XX

XX

PD WO9803632-A1.

XX

PD 29-JAN-1998.

XX

PF 23-JUL-1997; 97WO-US12911.

XX

PR 24-JUL-1996; 96US-0686934.

XX

XX

PA (UYA) UNIV YALE.

XX

PI Carithers MD, Lerner MR;

XX

DR WPI; 1998-120757/11.

XX

PT Bivalent agonist of G-protein coupled receptors containing two

PT ligand domains - bonded to molecular backbone, for treatment of

PT hypertension, promotion of skin tanning etc., also for delivering

PT drugs and gene therapy vectors to selected cells

XX

XX

PS Claim 41; Page 48; 71pp; English.

XX

CC This sequence represents one of two identical chains disulphide bonded

CC to form a bombesin dimer. The invention relates to bivalent agonists,

CC with affinity for at least 1 G-protein coupled receptor (GPCR). The

CC bivalent agonists comprise: (a) two ligand domains (LD), individually

CC agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)

CC a molecular backbone (MB) covalently bound to LD. The bivalent agonists

CC are useful in human or veterinary medicine as carriers for drugs or gene

CC therapy vectors, allowing these to be endocytosed by GPCR-expressing

CC cells. They can also be used e.g. to treat hypertension (angiotensin-

CC based LD); to increase levels of luteinising hormone (LH), using LD

CC derived from LH-releasing hormone, or to promote skin tanning (LD based

CC on alpha-melanocyte-stimulating hormone, MSH). The bivalent agonists

CC are administered orally, by injection or topically. Typical doses for

CC skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or

CC the bivalent agonists are administered topically in a composition

CC containing 0.001-10 (especially 1) mm. Where both LD are agonists, the

CC bivalent agonist has a synergistically higher activity than two

CC individual agonist ligands, and where at least 1 is an antagonist the

CC effect is stimulatory. The bivalent agonists are active at lower

CC concentrations than known agonists so should avoid toxicity problems.

XX

SO Sequence 39 AA:

Query Match 100.0%; Score 109; DB 19; Length 39;

Best Local Similarity 100.0%; Pred. No. 3.6e-06;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CXXXXXXXXXXXXXXXXX 21

DB 1 CXXXXXXXXXXXXXXXXX 21

RESULT 3

AAU18238

ID AAU18238 standard; Protein; 58 AA.

XX

AC AAU18238;

XX

DT 21-NOV-2001 (first entry)

XX

DE Novel human DNA-binding protein #85.

XX

KM Human: DNA-binding protein; histone; chromo domain protein;

KM chromatin organisation modifier; y-box binding protein;

KM DNA organisation; gene transcription; malignant disease;

KM autoimmune disorder; rheumatic disease; genetic abnormality;

KM infectious disease; neurological disorder; gene therapy;

KM immunomodulatory; anti-HIV; anti rheumatic; anti microbial;

KM cytosstatic.

XX

OS Homo sapiens.

XX

PN WO200155162-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01305.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216680.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. AA018154-AA018281 represent novel DNA-binding
CC proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 58 AA;

Query Match 100.0%; Score 109; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 5e-06; 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXX 21
|||||
DB 35 CXXXXXXXXXXXXXXXXX 55

RESULT 4
AA011210
ID AA011210 standard; Protein: 70 AA.
XX
XX AA011210;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 25102.
XX
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX N-PSDB; AA191141.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 70 AA;

Query Match 100.0%; Score 109; DB 22; Length 70;
Best Local Similarity 100.0%; Pred. No. 5.8e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXX 21
|||||
DB 30 CXXXXXXXXXXXXXXXXX 50

RESULT 5
AA003766
ID AA003766 standard; Protein: 81 AA.
XX
XX AA003766;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 17658.
XX
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX N-PSDB; AA183697.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 81 AA;

Query Match 100.0%; Score 109; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXX|
Db 48 CXXXXXXXXXXXXXXXXX 68

RESULT 6
AA011849
ID AA011849 standard; Protein; 36 AA.
XX
AC AA011849;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 25741.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR N-PSDB; AA191780.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 25741; 1399bp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 36 AA;

Query Match 96.3%; Score 105; DB 22; Length 36;
Best Local Similarity 95.2%; Pred. No. 9.6e-06;
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXX|
Db 11 CXXXXXXXXXXXXXXXXX 31

RESULT 7
AA012447

ID AA012447 standard; Protein; 47 AA.
XX
AC AA012447;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 26339.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR N-PSDB; AA192378.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 26339; 1399bp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 47 AA;

Query Match 96.3%; Score 105; DB 22; Length 47;
Best Local Similarity 95.2%; Pred. No. 1.2e-05;
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXX|
Db 12 CXXXXXXXXXXXXXXXXX 32

RESULT 8
AA018239
ID AA018239 standard; Protein; 66 AA.
XX
AC AA018239;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human DNA-binding protein #86.

XX
KW Human: DNA-binding protein; histone; chromo domain protein;
KW Chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic.
XX
OS Homo sapiens.
XX
PN W0200155162-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246510.
PR 08-NOV-2000; 2000US-0246511.
PR 08-NOV-2000; 2000US-0246513.
PR 17-NOV-2000; 2000US-0249257.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465557/50.
XX N-PSDB; AAC29115.
XX
PT Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
XX
XX Claim 11; SEQ ID No 224; 561pp; English.
XX
XX The present invention relates to the isolation of novel DNA-binding
XX proteins, and cDNA (AA529030-AA529157) and genomic sequences encoding
XX for these proteins. DNA-binding proteins such as histones, chromo-
XX (chromatin organization modifier) domain proteins, and Y-box binding
XX proteins may contribute to diseases resulting from aberrant DNA
XX organization and/or gene transcription. The sequences of the invention
XX are useful in screening assays to identify antagonists and/or agonists
XX that may enhance or block activities mediated by DNA-binding proteins.
XX Blockers of DNA-binding proteins may be useful in treating disorders
XX such as malignant diseases (e.g. cancer), autoimmune disorders
XX (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
XX arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
XX diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
XX disease). The polynucleotide sequences of the invention may also be
XX used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
XX proteins.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 66 AA:
SQ
Query Match 96.3%; Score 105; DB 22; Length 66;
Best Local Similarity 95.2%; Pred. No. 1.5e-05;
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 CCKKKKKKKKKKKKKKKKK 21
1:|||||
Db 38 cckkkkkkkkkkkkkkkkk 58
RESULT 9
AA53800
ID AAB53800 standard; Protein; 69 AA.
XX
XX AAB53800;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1340.
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX identification; cytostatic; cardioactive; neuroprotective; vulnery;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX neoplastic; antiinfective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;
XX infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX

PN W0200055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX N-PSDB; AAC38557.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX
XX Claim 11; Page 1920-1921; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AA53324 to AA54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnery, nephrotropic, antiinfective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies to the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins
XX may also be used to prevent diseases such as neural disorders, immune
XX system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, wounds, renal disorders, infectious
XX diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX AA54007 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 69 AA:
SQ
Query Match 96.3%; Score 105; DB 21; Length 69;
Best Local Similarity 95.2%; Pred. No. 1.6e-05;
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 CCKKKKKKKKKKKKKKKKK 21
1:|||||
Db 18 cckkkkkkkkkkkkkkkkk 38
RESULT 10
AAG73729
ID AAG73729 standard; Protein; 83 AA.
XX
XX AAG73729;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:4493.
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 14.
XX
XX Homo sapiens.
XX
XX W0200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX

PT disorders -
XX
PS Claim 20; SEQ ID NO 25106; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 68 AA;

Query Match 94.5%; Score 103; DB 22; Length 68;
Best Local Similarity 95.2%; Pred. No. 2.6e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 CXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXX
Db 17 cXXXXXXXXXXXXXXXXX 37

RESULT 13
AA000291
ID AA000291 standard; Protein; 124 AA.
XX
AC AA000291;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 14183.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PE 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR N-PSDB: AA180222.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 14183; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 124 AA;

Query Match 94.5%; Score 103; DB 22; Length 124;
Best Local Similarity 95.2%; Pred. No. 4.2e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 CXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXX
Db 20 cXXXXXXXXXXXXXXXXX 40

RESULT 14
AA012203
ID AA012203 standard; Protein; 41 AA.
XX
AC AA012203;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 26095.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PE 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR N-PSDB: AA192134.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 26095; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC Inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 41 AA;

Query Match 93.6%; Score 102; DB 22; Length 41;
Best Local Similarity 95.2%; Pred. No. 2.3e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CXXXXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXXXXX|
Db 4 CXXXXXXXXXXXXXXXXXXXX 24

RESULT 15

AAO12187

ID AAO12187 standard; Protein; 60 AA.

XX AAO12187;

XX

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 26079.

XX

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HKSE-) HKSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA192118.

PT Isolated nucleic acids and polypeptides, useful for preventing
PT disorders and treating e.g. leukaemia, inflammation and immune
PT disorders -

PS Claim 20; SEQ ID NO 26079; 1399pp + Sequence Listing: English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 60 AA;

Query Match 93.6%; Score 102; DB 22; Length 60;
Best Local Similarity 95.2%; Pred. No. 3.1e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CXXXXXXXXXXXXXXXXXXXX 21
|XXXXXXXXXXXXXXXXXXXX|
Db 26 CXXXXXXXXXXXXXXXXXXXX 46

Search completed: July 1, 2002, 06:19:22
Job time: 509 sec

Mon Jul 1, 13:54:41 2002

us-09-461-684-1.rag

Page 11

OY 2 EAAAAAEEAAAAAEEAAAAA 25
DB 115 QAEDAKAAAEAEKAAAAAEEAAAAA 138

RESULT 3
S19114
cgcr-1 protein - Chlamydomonas reinhardtii (fragment)
C:Species: Chlamydomonas reinhardtii
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 03-Nov-2000
C:Accession: S19114
R:Makarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A:Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A:Reference number: S19113; MUID:92119224
A:Accession: S19114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <WAK>
A:Cross-references: EMBL:X17207
C:Superfamily: phage lambda hypothetical protein 401

Query Match 65.1%; Score 71; DB 2; Length 205;
Best Local Similarity 75.0%; Pred. No. 0.91;
Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAAEEAAAAA 25
DB 49 EAAAAAQAATAAAAEAAARAPA 72

RESULT 4
S24960
gene C98 protein - rape
C:Species: Brassica napus (rape)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C:Accession: S24960
R:Roberts, M.R.; Hodge, R.; Sorensen, A.; Ross, J.; Murphy, D.J.; Draper, R.; Scott, R.
submitted to the EMBL data library, July 1992
A:Description: A new class of Brassica napus oleosin genes specific to the male gametoph
A:Reference number: S24960
A:Accession: S24960
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-183 <ROB>
A:Cross-references: EMBL:X67142; NID:g17792; PID:g17793

Query Match 64.2%; Score 70; DB 2; Length 183;
Best Local Similarity 78.3%; Pred. No. 1;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEEAAAAA 25
DB 151 AAPAAEPAPAAEAPAAAPAA 173

RESULT 5
A48423
engrailed homeodomain-containing protein En-1 - mouse
N:Alternate names: homeotic protein En-1
C:Species: Mus musculus (house mouse)
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C:Accession: A48423; S13009; A26629; A24778
R:Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallanathan, D.; Provart, N.J.; Joyner, A.I.
Dev. Genet. 13, 345-358, 1992
A:Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene
A:Reference number: A48423; MUID:93185339
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA

A:Residues: 1-401 <LOC>
A:Experimental source: CD-1, embryo
A:Note: sequence extracted from NCBI backbone (NCBIF:126620)
R:Holland, P.W.H.; Williams, N.A.
FEBS Lett. 277, 250-252, 1990
A:Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution
A:Reference number: S13009; MUID:91099509
A:Accession: S13009
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 321-380 <HOL>
R:Joyner, A.L.; Martin, G.R.
Genes Dev. 1, 29-38, 1987
A:Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engr
A:Reference number: A91620; MUID:88112776
A:Accession: A26629
A:Molecule type: DNA; mRNA
A:Residues: 278-401 <JOY>
A:Cross-references: GB:Y00201; GB:M11987; NID:549587; PIDN:CAA68361.1; PID:g669105
R:Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.
Cell 43, 29-37, 1985
A:Title: Expression during embryogenesis of a mouse gene with sequence homology to t
A:Reference number: A24778; MUID:86079501
A:Accession: A24778
A:Molecule type: DNA
A:Residues: 311-401 <JO2>
C:Genetics:
A:Gene: en-1
A:Map position: 1
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:313-369/Domain: homeobox homology <HOX>

Query Match 64.2%; Score 70; DB 2; Length 401;
Best Local Similarity 78.3%; Pred. No. 1.9;
Matches 18; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAEEAAAAA 25
DB 207 AAAAAAATAAAAVAAAAAAS 229

RESULT 6
S02376
antifreeze protein precursor - yellowtail flounder
C:Species: Limanda ferruginea (yellowtail flounder)
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 24-Oct-2000
C:Accession: S02376
R:Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.
Eur. J. Biochem. 168, 629-633, 1987
A:Title: Structural variations in the alanine-rich antifreeze proteins of the Pleuro
A:Reference number: S02376; MUID:88029483
A:Accession: S02376
A:Molecule type: mRNA
A:Residues: 1-97 <SCO>
A:Cross-references: EMBL:X06356; NID:g64041; PIDN:CAA29655.1; PID:g64042
A:Note: part of this sequence, including the amino end of the mature protein, was co
C:Superfamily: antifreeze protein
C:Keywords: antifreeze
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-48/Domain: propeptide #status predicted <PRO>
F:49-96/Product: antifreeze protein #status predicted <MAT>

Query Match 63.3%; Score 69; DB 2; Length 97;
Best Local Similarity 70.8%; Pred. No. 0.82;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAEEAAAAAEEAAAAA 25
DB 53 DAAAAAATAAAAKAAADTAAAA 76

[illegible]

```

RESULT 12
533225
ecdysone-induced protein E74A - fruit fly (Drosophila virilis)
C:Species: Drosophila virilis
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Feb-1997
C:Accession: B53225
R:Jones, C.W.; Dalton, M.W.; Townley, L.H.
Genetics 127, 535-543, 1991
A:Title: Interspecific comparisons of the structure and regulation of the Drosophila ecd
A:Reference number: A53225; MUID:91200627
A:Accession: B53225
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-873 <GR>
A:Cross-references: GB:X59493
C:Genetics:
A:Gene: FlyBase:Dftr/Efp74E
A:Cross-references: FlyBase:FBgn0013076
C:Superfamily: ets DNA-binding domain homology
:779-859/domain: ets DNA-binding domain homology <ETS>

```

RESULT 13
T13154
polycomb protein enhancer - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13154
R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.
Development 125, 4055-4066, 1998
A:Title: The enhancer of polycomb gene of *Drosophila* encodes a chromatin protein conserv
A:Reference number: 217611; MUID:98407961
A:Accession: T13154
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2023 <STA>
A:Cross-references: EMBL:AF079764; NID:g3757889; PID:g3757890; FIDN:AA64271.1
A:Experimental source: imaginal disc
A:Genetics:
A:Gene: E(Pc)
A:Cross-references: flybase:FBgn0000561
A:Map position: 2

RESULT 14
JC5273
paired type homeobox protein, NBP - human
C:Species: Homo sapiens (man)
C:Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 24-Sep-1999

C:Accession: Jc5273
R:Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.
DNA Res. 3, 311-320, 1996
A>Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific
A:Reference number: Jc5272; MUID:97191543
A:Contents: neuroblastoma cell
A:Accession: Jc5273
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-314 <OK>
A:Cross-references: DBPJ:D82344; NID:q1841337; PIDN:BA41155.1; PID:d1012222; PID:q1841337
A:Comment: This protein is a transcriptional repressor involved in regulating gene expression
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:95-15/Domain: homeobox homology <Hox>

RESULT 15
T33110
Hypothetical protein C18H7.3 - *Caenorhabditis elegans*
C|Species: *Caenorhabditis elegans*
C|Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C|Accession: T33110
R|Tin-Mollam, A.; Fronick, W.
Submitted to the EMBL Data Library, May 1998
A|Description: The sequence of C. elegans cosmid C18H7.
A|Reference number: 221284
A|Accession: T33110
A|Status: preliminary; translated from GB/EMBL/DBD
A|Molecule type: DNA
A|Residues: 1-460 <IN>
A|Cross-references: EMBL:AF067607; PIDN:AA017641.1; GSPDB:GN00022; CESP:C18H7.3
A|Experimental source: strain Bristol N2; clone C18H7
C|Genetics:
A|Gene: CESP:C18H7.3
A|Map position: 4
A|Intons: 84/1
A|Superfamily: Phaseolus glycinine-rich cell wall protein 1.8

Search completed: July 1, 2002, 06:28:10
Job time: 712 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:33 ; Search time 25.37 Seconds
(without alignments)
38.155 Million cell updates/sec

Title: US-09-461-684-2
Perfect score: 109
Sequence: 1 CEAAAAAEPAFAAAAAEPAFAAAAA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.5	69.3	518	1	TPM4_DROME
2	72	66.1	185	1	RS16_CAUCR
3	70	64.2	183	1	OLEC_BRANA
4	70	64.2	401	1	HME1_MOUSE
5	69	63.3	97	1	ANP_LIME
6	69	63.3	1028	1	ANO_DROME
7	68	62.4	673	1	HR38_DROME
8	68	62.4	1073	1	HR38_DROME
9	67	61.5	314	1	PMXR_MOUSE
10	67	61.5	314	1	PMXR_MOUSE
11	67	61.5	477	1	MA2_HUMAN
12	67	61.5	477	1	MA2_HUMAN
13	66	60.6	366	1	HMD_DROAN
14	66	60.6	388	1	HMD_DROAN
15	65	59.6	85	1	HMD_MOUSE
16	65	59.6	85	1	HMD_MOUSE
17	65	59.6	91	1	ANPX_PSEAM
18	65	59.6	276	1	ANPX_PSEAM
19	65	59.6	475	1	EVX2_MOUSE
20	65	59.6	476	1	EVX2_MOUSE
21	64	58.7	644	1	BTD_DROME
22	64	58.7	109	1	RLAI_TRYCR
23	64	58.7	392	1	HME1_HUMAN
24	64	58.7	1533	1	HME1_HUMAN
25	64	58.7	2038	1	HME1_HUMAN
26	63	57.8	376	1	FXE1_HUMAN
27	63	57.8	521	1	FXE1_HUMAN
28	63	57.8	590	1	HMA2_DROME
29	63	57.8	607	1	HMA2_DROME
30	63	57.8	1095	1	PIPA_MOUSE
31	62	56.9	364	1	PIPA_MOUSE
32	62	56.9	365	1	PIPA_MOUSE
33	62	56.9	1355	1	PIPA_MOUSE
34	61	56.0	31	1	ANP3_PAGBO

ALIGNMENTS

RESULT	ID	TPM4_DROME	STANDARD	PRT	518 AA
AC	TPM4_DROME	TPM4_DROME	STANDARD	PRT	518 AA
AC	P49455	P49455			
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DE	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Tropomyosin I, fusion protein 33.				
GN	TMI OR TMI.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=89127197; PubMed=2851721;				
BA	Hanke P.D., Scott R.V.				
RT	"The drosophila melanogaster tropomyosin II gene produces multiple proteins by use of alternative tissue-specific promoters and alternative splicing."				
RT	Mol. Cell. Biol. 8:3591-3602(1988).				
RN	[2]				
RP	SEQUENCE OF 1-286 FROM N.A.				
RP	MEDLINE=87064486; PubMed=3097506;				
BA	Karlik C.C., Pyberg E.A.				
RT	"Two Drosophila melanogaster tropomyosin genes: structural and functional aspects."				
RT	Mol. Cell. Biol. 6:1965-1973(1986).				
CC	-1- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN I GENE CAN PRODUCE FOUR DIFFERENT ISOFORMS BY ALTERNATIVE PRODUCTS: A MUSCLE FORM, A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).				
CC	-1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.				
CC	-1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.				
CC	-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE OF ALTERNATIVE EXON USAGE.				
CC	*****				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb.slb.ch).				
CC	*****				
DR	EMBL: X76208; CA53800.1; -				P23699 pseudopleur
DR	EMBL: L00362; AAA28965.1; -				P39413 drosophila
DR	EMBL: M12840; AAA28965.1; JOINED.				P58012 homo sapien
DR	EMBL: L00335; AAA28965.1; JOINED.				P17208 mus musculu
DR	EMBL: L00336; AAA28965.1; JOINED.				P01851 homo sapien
DR	EMBL: L00337; AAA28965.1; JOINED.				P06654 streptococ
DR	EMBL: L00338; AAA28965.1; JOINED.				P19909 streptococ
DR	EMBL: L00339; AAA28965.1; JOINED.				P33485 pseudorabie
DR	HSSP: P04002; IATF.				O18881 macaca radi
DR	Flybase: FBgn0003721; Tml.				P04002 pseudopleur
DR					P16527 gallus gall
DR					P78426 homo sapien

```

DR InterPro: IPR000533; Tropomyosin.
DR Pfam: PF00261; Tropomyosin; 1.
DR PRINTS: PR00194; TROPOMYOSIN.
DR PROSITE: PS00326; TROPOMYOSIN; 1.
FT COILED COIL; Repeat; Alternative splicing; Multigene family.
FT DOMAIN 14 267 COILED COIL (POTENTIAL).
FT DOMAIN 287 518 ALA/PRO-RICH.
FT CONFLICT 106 114 LGSATKLS -> SASAIQLAA (IN REF. 2).
FT CONFLICT 119 119 A -> S (IN REF. 2).
FT CONFLICT 183 183 A -> AMVEADLERASERA (IN REF. 2).
FT CONFLICT 199 199 V -> L (IN REF. 2).
FT CONFLICT 231 231 MOREEKKYKQIKTLNTR -> TQKEEFETQIKVLDS
(IN REF. 2).
SQ SEQUENCE 518 AA; 54558 MW; 153D0872CF9DB6EA CRC64;

Query Match 69.3%; Score 75.5; DB 1; Length 518;
Best Local Similarity 87.3%; Pred. No. 0.2;
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAAAAAAAAAAAAA 25
DB 314 EAAAAEAAAAAAAAAEAAAAA 336

RESULT 2
RS16_CAUCR STANDARD; PRT; 165 AA.
AC PS8122;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 30S ribosomal protein S16.
GN RSP OR CC3652.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
CC Caulobacter
NCBI_TaxID:63394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE-21173698; PubMed-11259647;
RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Madcock J.R.,
RA Potocky I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Usterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1 SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AEO06023; AAK25614.1; -
DR TIGR: CG3652; -
DR InterPro: IPR000307; Ribosomal_S16.
DR Pfam: PF00886; Ribosomal_S16; 1.
DR PROSITE: PS00732; RIBOSOMAL_S16; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 165 AA; 17605 MW; ED46FC2798C5BE1C CRC64;

Query Match 66.1%; Score 72; DB 1; Length 165;
Best Local Similarity 75.0%; Pred. No. 0.18;
Matches 18; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 2 EAAAAEAAAAAAAAAAAAAAAA 25
DB 115 CAEADAKAAAEAKKAAAEAAAA 138

RESULT 3
OLEC_BRANA STANDARD; PRT; 183 AA.
AC P29526;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Oleosin C98 (Fragment).
GN C98
OS Brassica napus (rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Anther;
RX MEDLINE-93386188; PubMed-8374615;
RA Roberts M.R., Hodge R., Ross J.H.E., Sorensen A., Murphy D.J.,
RA Draper J., Scott R.;
RT "Characterization of a new class of oleosins suggests a male
RT gametophyte-specific lipid storage pathway.";
RT Plant J. 3:628-636(1993).
CC -1 FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
CC DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE
CC OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOETIES
CC OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
CC LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
CC -1 SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
CC MONOLAYER LIPID/WATER INTERFACE.
CC -1 TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.
CC -1 SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X67142; CAA47623.1; -
DR PIR: S24960; S24960.
DR HSP: P04002; 1MFA.
DR InterPro: IPR000136; Oleosin.
DR Pfam: PF01277; Oleosin; 1.
DR PROSITE: PS00811; OLEOSINS; 1.
KW Seed; Oil body; Multigene family.
FT NON_TER 1 1
FT DOMAIN <1 23 POLAR.
FT DOMAIN 24 95 HYDROPHOBIC.
SQ SEQUENCE 183 AA; 18149 MW; 198A5D3BD7F3045A CRC64;

Query Match 64.2%; Score 70; DB 1; Length 183;
Best Local Similarity 78.3%; Pred. No. 0.31;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAAAAEAAAAAAAAAAAAAAAA 25
DB 151 AAPAAEPAPAAEAPAAEAPAPAA 173

RESULT 4
HME1_MOUSE STANDARD; PRT; 401 AA.
ID HME1_MOUSE
AC P09065;

```


16-OCT-2001 (Rel. 40, last annotation update)
 DE Ovo protein (Shaven baby protein).
 GN OVO OR SVB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Meyel-Nitlo M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development.";
 RL EMBO J. 10:2259-2266(1991).
 CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMIUM AND
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U11383; AAB60216.1; -;
 DR EMBL: X59772; CAB36921.1; ALT_SEQ.
 DR HSSP: P25490; 12NM.
 DR Flybase; FBgn0003028; ovo.
 DR InterPro; IPR000822; znf-C2H2.
 DR Pfam; PF00096; znf-C2H2; 4.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00355; znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 539 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.

FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 826 832 POLY-GLN.
 FT DOMAIN 874 992 ZINC_FINGERS.
 FT ZN_FING 874 896 C2H2-TYPE.
 FT ZN_FING 902 924 C2H2-TYPE.
 FT ZN_FING 930 953 C2H2-TYPE.
 FT ZN_FING 969 992 C2H2-TYPE.
 FT CONFLICT 647 647 A -> R (IN REF. 2).
 SQ SEQUENCE 1028 AA; 110620 MW; D7068B2BC0F6F77 CRC64;
 Query Match 63.3%; Score 69; DB 1; Length 1028;
 Best Local Similarity 85.7%; Pred. No. 1.5;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 AAAAAAAAAAAAAAAAAAAAAA 23
 Db 497 AAAAAAAAAAAAAAAAAAAAAA 517
 SIM_DROME 7
 ID SIM_DROME STANDARD; PRT; 673 AA.
 AC P05709; O96521; Q9VF23;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Single-minded protein.
 GN SIM OR CG7771.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_TaxID=7227;
 RN [1]
 RP SEQUENCE OF 19-673 FROM N.A.
 RX MEDLINE=88151023; PubMed=3345560;
 RA Crews S.T., Thomas J.B., Goodman C.S.;
 RT "The Drosophila single-minded gene encodes a nuclear protein with
 RT sequence similarity to the per gene product";
 RL Cell 52:143-151(1988).
 RN [2]
 RP SEQUENCE OF 1-18 FROM N.A. AND SIMILARITY TO HLH PROTEINS.
 RX MEDLINE=92103681; PubMed=1760843;
 RA Namdu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;
 RT "The Drosophila single-minded gene encodes a helix-loop-helix protein
 RT that acts as a master regulator of CNS midline development.";
 RL Cell 67:1157-1167(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99054545; PubMed=9840810;
 RA Kasal Y., Stahl S., Crews S.;
 RT "Specification of the Drosophila CNS midline cell lineage: direct
 RT control of single-minded transcription by dorsal/ventral patterning
 RT genes";
 RL Gene Expr. 7:171-189(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Adney A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bandaru D., Boltskov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
 RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foser C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Krevitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sigen Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RL "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER
 CC DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN
 CC THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING
 CC RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS.
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: M19020; AAC28900.1; -;
 DR EMBL: AF071934; AAC64519.1; ALT_SEQ.
 DR EMBL: AE003698; AAF54902.1; ALT_SEQ.
 DR PIR: A29945; A29945.
 DR TRANSFAC: T00750; -;
 DR FLYBASE: FBgn0004666; sim.
 DR InterPro: IPR003015; HLH_MYC.
 DR InterPro: IPR001092; HLH_dlm.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF00785; PAC; 1.
 DR Pfam: PF00989; PAS; 2.
 DR SMART: SM00353; HLH; 1.
 DR SMART: SM00086; PAC; 1.
 DR SMART: SM00091; PAS; 2.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
 DR PROSITE: PS50112; PAS; 2.
 KW Developmental protein; Neurogenesis; Nuclear protein; Repeat;
 KW Transcription regulation; DNA-binding
 FT DNA_BIND 1 13
 FT DOMAIN 14 54 BASIC DOMAIN
 FT DOMAIN 54 148 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 148 148 PAS 1.
 FT DOMAIN 242 312 PAS 2.
 FT DOMAIN 312 422 14 x 3 AA REPEATS OF A-A-Q.
 FT DOMAIN 422 426 POLY-VAL.
 FT DOMAIN 426 488 POLY-SER.
 FT DOMAIN 488 488 POLY-ASN.
 FT DOMAIN 489 492 POLY-HIS.
 FT DOMAIN 492 496 POLY-GLN.
 FT DOMAIN 501 505

FT DOMAIN 529 533 POLY-SER.
 FT DOMAIN 573 576 POLY-ASN.
 FT DOMAIN 604 607 POLY-SER.
 FT DOMAIN 649 669 GLN/HIS-RICH.
 FT CONFLICT 127 127 I -> Y (IN REF. 3).
 FT CONFLICT 401 409 MISSING (IN REF. 4).
 SQ SEQUENCE 673 AA; 73589 MW; 29F0ABBA2BC0CBE CRC64;
 Query Match 62.4%; Score 68; DB 1; Length 673;
 Best Local Similarity 66.7%; Pred. No. 1.3;
 Matches 16; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Cy 2 EAAAEAAAEAAAEAAAEAAAEAA 25
 Db 384 QAAQAQAQAQAQAQAQAQAQA 407
 RESULT 8
 ID HR38_DROME STANDARD; PRT; 1073 AA.
 AC PA9858: O18383; OSYITKA;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Probable nuclear hormone receptor HR38 (dHR38).
 GN HR38 OR NR4A4 OR CG1864.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE-Larva;
 RX MEDLINE=95372400; PubMed=7644522;
 RA Sutherland J.D., Kozlova T., Tzertzinis G., Kafatos F.C.;
 RT "Drosophila hormone receptor 38: a second partner for *Drosophila* USP
 RT factor-induced protein B type.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7966-7970(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE SPECIFICITY.
 RX MEDLINE=98370123; PubMed=9704500;
 RA Komonyi O., Mink W., Csiba U., Maroy P.;
 RT "Genomic organization of DHR38 gene in *Drosophila*: presence of
 RT Alu-like repeat in a translated exon and expression during embryonic
 RT development.";
 RL Arch. Insect Biochem. Physiol. 38:185-192(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG ISOFORM), AND CHARACTERIZATION.
 RC TISSUE-Larva;
 RX MEDLINE=98315108; PubMed=9649534;
 RA Kozlova T., Pokholkova G.V., Tzertzinis G., Sutherland J.D.,
 RA Zhimulev I.F., Kafatos F.C.;
 RT "Drosophila hormone receptor 38 functions in metamorphosis: a role in
 RT adult cuticle formation.";
 RL Genetics 149:1465-1475(1998).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner R.A., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abill J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bayendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brocksen P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabriellian A.E., Garg N.S., Gebhart W.M., Glasser K.,
 Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 RL [5]
 RT
 RN SEQUENCE OF 528-1073 FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE=9606864; PubMed=7479849;
 RA Flisk G.J., Thummel C.S.;
 RT "Isolation, regulation, and DNA-binding properties of three
Drosophila nuclear hormone receptor superfamily members.";
 Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608(1995)
 RL
 CC -1- FUNCTION: BINDS TO NGF-R RESPONSE ELEMENTS. PLAYS AN IMPORTANT
 ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.
 CC
 CC -1- SUBUNIT: FORMS A HETERODIMER WITH USP.
 CC
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC
 CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN PREBLASTODERM
 EMBRYOS SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL
 TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND
 BRAIN COMPLEXES, BUT NOT IN OVARIES.
 CC
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.
 CC
 CC HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.
 CC
 CC SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN
 LARVAE.
 CC
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC
 CC N4 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL: X89246; CAA61534.1; -
 DR EMBL: Y15606; CAA75690.1; -
 DR EMBL: AJ002073; CAA05172.1; -
 DR EMBL: AE003667; AAF53914.1; -
 DR EMBL: U36762; AAC46926.1; -
 DR HSSP: P19793; NULL.
 DR TRANSFAC: T02760; -
 DR FlyBase: FBgn0014859; Hx38.
 DR InterPro: IPR000536; Hormone_rec_119.
 DR InterPro: IPR001723; StrdHormone_receptor.
 DR InterPro: IPR001628; zf-C4.
 DR Pfam: PF00104; hormone_rec_2.
 DR Pfam: PF00105; zf-C4; 2.

DR PRINTS; PR00398; STRDHOMONER.
 DR PRINTS; PR00447; STROIDFINER.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00399; zNF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger; Alternative splicing; Developmental protein.
 FT
 FT DNA_BIND 744 809
 FT ZN_FING 744 764
 FT ZN_FING 780 804
 FT ZN_FING 802 824
 FT DOMAIN 188 192
 FT DOMAIN 206 218
 FT DOMAIN 221 228
 FT DOMAIN 221 228
 FT DOMAIN 258 272
 FT DOMAIN 294 312
 FT DOMAIN 441 462
 FT DOMAIN 505 508
 FT DOMAIN 619 626
 FT DOMAIN 651 655
 FT VARSPLIC 1 522
 FT CONFLICT 667 667
 FT CONFLICT 667 685
 FT CONFLICT 685 692
 FT CONFLICT 692 697
 FT CONFLICT 697 702
 FT CONFLICT 702 702
 FT CONFLICT 1041 1041
 FT CONFLICT 1064 1064
 SQ SEQUENCE 1073 AA; 116991 MW; 126A30DAFAC096A CRC64;

 Query Match 62.4%; Score 68; DB 1; Length 1073;
 Best Local Similarity 73.9%; Pred. No. 1.9;
 Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

 QY 3 AAAAAAATAAATAAATAAATAAATAA 25
 Db 206 AATAAATAAATAAATAAATAAATAA 228

 RESULT 9
 ID PMXB_HUMAN STANDARD; PRT; 314 AA.
 AC 099453;
 DT 30-MAY-2000 (rel. 39, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
 DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBphox).
 GN PMXB2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Neuroblastoma;
 RX MEDLINE=97191543; PubMed=9039501;
 RA Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;
 RT Identification and cloning of neuroblastoma-specific and nerve
 RT tissue-specific genes through compiled expression profiles.";
 RL DNA Res. 3:311-320(1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20487360; PubMed=11034547;
 RA Adachi M., Browne D., Lewis E.J.;
 RT "Paired-like homeodomain proteins Phox2a/Atx and Phox2b/NBphox have
 RT similar genetic organization and independently regulate dopamine
 RT beta-hydroxylase gene transcription.";
 RL DNA Cell Biol. 19:539-554(2000).
 RL [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99326521; PubMed=10395798;
 RA Yokoyama M., Watanabe H., Nakamura M.;

RT "Genomic structure and functional characterization of NBPHox (PMX2B),
RT a homeodomain protein specific to catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation."
RL Genomics 59:40-50(1999)
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR
CC NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.
CC TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEURORANSAMETER
CC PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED
CC ACTIVATION OF THE DOPAMINE BETA-HYDROLYASE AND C-FOS PROMOTERS,
CC AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND
CC SERUM-RESPONSE ELEMENT
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL
CC GLAND.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D82344; BA01555.1; -
DR EMBL: AF117979; A026698.1; -
DR EMBL: AB015671; BA082670.1; -
DR HSSP: P14653; 1B72.
DR MIM: 603851; -
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA_BIND 98 157 HOMEBOX.
FT DOMAIN 159 167 POLY-ALA.
FT DOMAIN 212 217 POLY-GLY.
FT DOMAIN 241 260 POLY-ALA.
SQ SEQUENCE 314 AA; 31607 MW; 76737F71948B5D81 CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 0.95;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAEAAAAEAAAAA 25
Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266

RESULT 10
PMXB_MOUSE STANDARD; PRT; 314 AA.
AC 035690;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN PMX2B OR PHOX2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98040559; PubMed=9374403;
RA Patten A., Morin X., Cremer H., Goridis C., Brunet J.-F.;
RT "Expression and interactions of the two closely related homeobox
RT genes Phox2a and Phox2b during neurogenesis.";

RL Development 124:4065-4075(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99326521; PubMed=10395798;
RA Yokoyama M., Watanabe H., Nakamura M.;
RT "Genomic structure and functional characterization of NBPHox (PMX2B),
RT a homeodomain protein specific to catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation."
RL Genomics 59:40-50(1999).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y14493; CA07483.1; -
DR EMBL: AB015672; BA082671.1; -
DR HSSP: P06601; 1F3L.
DR MGD: MGI:1100882; Pmx2b.
DR InterPro: IPR00047; HTH_repressr.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR00031; HTHREPRESSR.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
KW Transcription regulation.
FT DNA_BIND 98 157 HOMEBOX.
FT DOMAIN 159 167 POLY-ALA.
FT DOMAIN 212 217 POLY-GLY.
FT DOMAIN 241 260 POLY-ALA.
SQ SEQUENCE 314 AA; 31621 MW; 40737F71948B595A CRC64;

Query Match 61.5%; Score 67; DB 1; Length 314;
Best Local Similarity 78.3%; Pred. No. 0.95;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAEAAAAEAAAAEAAAAA 25
Db 244 AAAAAAAAAAAAAAAAAAGLAAA 266

RESULT 11
MAZ_HUMAN STANDARD; PRT; 477 AA.
AC P56270; Q99443; Q15703;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MYC-associated zinc finger protein (MAZ1) (Purine-binding
DE transcription factor) (Pur-1) (Zf87) (Zf87).
GN MAZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92366479; PubMed=1502157;
RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
RT regulating transcriptional initiation and termination."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Carcinoma;

RX MEDLINE=9232709; PubMed=1567856;
 RA Pyro J.J., Moberg K.H., Hall D.J.;
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds
 to two sites within the c-myc promoter.";
 RL Biochemistry 31:4102-4110(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreatic Islets;
 RA MEDLINE=96428591; PubMed=8931693;
 RA Tsuboi H., Sakatsume O., Itakura K., Yokoyama K.K.;
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human
 pancreatic islet cells.";
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96224025; PubMed=8626793;
 RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 responds to MAZ and Sp1.";
 RL J. Biol. Chem. 271:4417-4430(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoblastoma;
 RA MEDLINE=96352105; PubMed=9685418;
 RA Song J., Murakami H., Tsuboi H., Tang X., Matsumura M., Itakura K.,
 Kanazawa I., Sun K., Yokoyama K.K.;
 RT "Genomic organization and expression of a human gene for Myc-
 associated zinc finger protein (MAZ).";
 RL J. Biol. Chem. 273:20603-20614(1998).
 CC -1- FUNCTION: MAZ FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
 IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
 MEAL AND KEAL, WITHIN THE C-MYC PROMOTER HAVING GREATER
 AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
 WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
 MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M94046; -; NOT_ANNOTATED_CDS.
 DR EMBL; M93339; -; NOT_ANNOTATED_CDS.
 DR EMBL; D85131; BA012728.1; ALT_INIT.
 DR EMBL; U38139; AAB04121.1; ALT_INIT.
 DR EMBL; AB017335; BAA33064.1; -;
 DR HSSP; P08046; IAAV;
 DR TRANSFAC; T00490; -;
 DR TRANSFAC; T02305; -;
 DR MIM; 600999; -;
 DR InterPro; IPR000822; Znf-C2H2.
 DR Pfam; PF00096; zf-C2H2; 6.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00355; Znf_C2H2; 6.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
 KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
 RNA-binding; Repeat; Nuclear protein.
 KW DOMAIN 190 413 ZINC_FINGERS.
 FT ZN_FING 190 212 C2H2-TYPE.
 FT ZN_FING 279 301 C2H2-TYPE.
 FT ZN_FING 307 329 C2H2-TYPE.
 FT ZN_FING 337 360 C2H2-TYPE.
 FT ZN_FING 366 388 C2H2-TYPE.
 FT ZN_FING 392 413 C2H2-TYPE.
 FT ZN_FING 96 108 POLY-ALA.
 FT DOMAIN 133 139 POLY-PRO.
 FT DOMAIN 157 161 POLY-ALA.

FT DOMAIN 245 249 POLY-GLY.
 FT DOMAIN 435 449 POLY-ALA.
 FT CONFLICT 259 259 MISSING (IN REF. 3).
 FT CONFLICT 401 401 L->M (IN REF. 2 AND 4).
 FT CONFLICT 443 447 MISSING (IN REF. 3).
 SQ SEQUENCE 477 AA; 48607 MW; C04C08F32C36825 CnC64;

Query Match 61.5%; Score 67; DB 1; Length 477;
 Best Local Similarity 77.3%; Pred. No. 1.3;
 Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 AAAAFAAAAAFAAAAAFAAAAA 25
 DB 90 AAAAFAAAAAFAAAAAFAAAAA 111

RESULT 12
 HMID_DROAN STANDARD; PRT; 606 AA.
 AC P22544;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Homeobox protein OM(1D).
 GN OM(1D).
 OS Drosophila ananassae (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxId=7217;
 RX MEDLINE=9112048; PubMed=1671353;
 RA Tanga S., Cores V.G.;
 RT "Retrotransposon-induced overexpression of a homeobox gene causes
 defects in eye morphogenesis in Drosophila.";
 RL EMO J. 10:407-417(1991).
 CC -1- FUNCTION: Probably involved in eye morphogenesis.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEOBOX PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X56682; CAA40011.1; -;
 DR PIR; S13367; S13367.
 DR HSSP; P22808; 1YND.
 DR Flybase; FBgn0012114; Dana/B.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00024; HOMEOBOX.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEOBOX_1; 1.
 DR PROSITE; PS50071; HOMEOBOX_2; 1.
 KW DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.
 KW DOMAIN 23 57 HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 106 124 HIS/GLN-RICH (OPA-REPEAT).
 FT DOMAIN 173 193 HIS/PRO-RICH.
 FT DNA_BIND 331 390 HOMEOBOX.
 FT DOMAIN 220 248 ALA-RICH.
 FT DOMAIN 422 434 ALA-RICH.
 FT DOMAIN 450 455 ALA-RICH.
 FT DOMAIN 503 510 ALA-RICH.
 FT DOMAIN 515 521 PRO-RICH.
 SO SEQUENCE 606 AA; 61735 MW; AA7B8B6367370FBB CnC64;

Query Match 61.5%; Score 67; DB 1; Length 606;
 Best Local Similarity 78.3%; Pred. No. 1.6;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AAAAAAAAAAAAAAAAAAAAAA 25
 DB 220 AAAAAAAAAAAAAAAAAAAAAA 242

RESULT 13
 HXAD_MOUSE STANDARD; PRT; 386 AA.
 AC Q6242;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Homeobox protein Hox-A13 (Hox-1.10).
 GN HOXA13 OR HOX-1.10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96259555; PubMed=8673126;
 RA Mortlock D.P., Post L.C., Innis J.W.;
 RT "The molecular basis of hypodactyly (hd): a deletion in Hoxa 13 leads
 RT to arrest of digital arch formation.";
 RL Nat. Genet. 13:284-289(1996).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A
 CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH
 CC STRUCTURES.
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U59322; AAB03322.1; -
 CC DR HSSP: P02833; ISAN.
 CC DR TRANSFAC: T03337; -
 CC DR MGD: MGI:96173; Hoxa13.
 CC DR InterPro: IPR001356; Homeobox.
 CC DR Pfam: PF00046; homeobox; 1.
 CC DR SMART: SM00389; HOX; 1.
 CC DR PROSITE: PS00027; HOMEBOX_1; 1.
 CC DR PROSITE: PS50071; HOMEBOX_2; 1.
 CC DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 CC Transcription regulation.
 CC KM Transcription regulation.
 CC FT DOMAIN 38 51 POLY-ALA.
 CC FT DNA_BIND 320 379 HOMEBOX.
 CC FT DOMAIN 52 57 POLY-GLY.
 CC FT DOMAIN 62 66 POLY-ALA.
 CC FT DOMAIN 73 84 POLY-ALA.
 CC FT DOMAIN 101 104 POLY-ALA.
 CC FT DOMAIN 116 133 POLY-ALA.
 CC FT DOMAIN 198 205 POLY-ALA.
 CC SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

Query Match 60.6%; Score 66; DB 1; Length 386;
 Best Local Similarity 70.8%; Pred. No. 1.4;
 Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 111 EAPPSAAAAAAAAAAAAAAAAAAS 134

RESULT 14
 HXAD_HUMAN STANDARD; PRT; 388 AA.
 AC P31271; O43371;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-A13 (Hox-1J).
 GN HOXA13 OR HOX1J.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97172976; PubMed=9020844;
 RA Mortlock D.P., Innis J.W.;
 RT "Mutation of HOXA13 in hand-foot-genital syndrome.";
 RL Nat. Genet. 15:179-180(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bradshaw H., Hinds K., Keppeler D.;
 RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 322-387 FROM N.A.
 RX MEDLINE=90098676; PubMed=2574852;
 RA Acampora D., D'Esposito M., Falcetta A., Panese M., Migliaccio E.,
 RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
 RT "The human Hox gene family.";
 RL Nucleic Acids Res. 17:10385-10402(1989).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U82827; AAC50993.1; -
 CC DR EMBL: AC004080; -; NOT_ANNOTATED_CDS.
 CC DR PIR: S14932; S14932.
 CC DR HSSP: P02833; ISAN.
 CC DR TRANSFAC: T03321; -
 CC DR MIM: 142959; -
 CC DR InterPro: IPR001356; Homeobox.
 CC DR Pfam: PF00046; homeobox; 1.
 CC DR SMART: SM00389; HOX; 1.
 CC DR PROSITE: PS00027; HOMEBOX_1; 1.
 CC DR PROSITE: PS50071; HOMEBOX_2; 1.
 CC DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 CC Transcription regulation.
 CC KM Transcription regulation.
 CC FT DOMAIN 38 53 POLY-ALA.
 CC FT DNA_BIND 322 381 HOMEBOX.
 CC FT DOMAIN 62 66 POLY-ALA.
 CC FT DOMAIN 73 84 POLY-ALA.
 CC FT DOMAIN 116 133 POLY-ALA.
 CC FT DOMAIN 145 150 POLY-ALA.
 CC FT DOMAIN 200 207 POLY-ALA.
 CC FT CONFLICT 146 146 A -> G (IN REF. 2).
 CC FT CONFLICT 147 147 A -> H (IN REF. 2).
 CC FT CONFLICT 187 187 P -> A (IN REF. 2).
 CC FT CONFLICT 195 195 P -> A (IN REF. 2).
 CC FT CONFLICT 198 198 P -> A (IN REF. 2).

SEQ SEQUENCE 388 AA; 39752 MW; 6CD9C9A5616C2FF6 CRC64;

Query Match 60.6%; Score 66; DB 1; Length 388;
Best Local Similarity 70.8%; Pred. No. 1.4;
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
|| : ||||| ||||| |||||
DB 111 EAPPSAAAAAAAAAAAAAAAAAS 134

RESULT 15

ANP4_PSEAM STANDARD; PRT; 85 AA.
ID ANP4_PSEAM
AC P02734; 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antifreeze peptide 4 precursor.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes; CC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81247379; PubMed=6265915;
RA Lin Y., Gross J.K.;
RT "Molecular cloning and characterization of winter flounder antifreeze RT CDNA";
RL Proc. Natl. Acad. Sci. U.S.A. 78:2825-2829(1981).
CC -! FUNCTION: ANTIFREEZE PROTEIN LOWER THE BLOOD FREEZING POINT.
CC -! SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; J00930; AAA49467.1; -
DR PIR; A03193; FDELAW.
DR HSSP; P04002; IWEA.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
KW Antifreeze protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 85 ANTIFREEZE PEPTIDE 4.
FT SEQUENCE 85 AA; 7215 MW; 8E62E1D2B44117BC CRC64;
SQ

Query Match 59.6%; Score 65; DB 1; Length 85;
Best Local Similarity 73.9%; Pred. No. 0.55;
Matches 17; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
||||| ||||| ||||| |||||
DB 28 AAAAAATATAAAAAAAAAATATA 50

Search completed: July 1, 2002, 06:31:34
Job time: 671 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:54 ; Search time 83.98 Seconds

(Without alignments)
51.499 Million cell updates/sec

Title: US-09-461-684-2
Perfect score: 109
Sequence: 1 CEAATAAATAAATAAATAAATAAATAA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL.19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeoprotein:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	77.1	2451	5	Q9VG05
2	80	73.4	1354	11	Q9EPW8
3	79	72.5	110	11	Q9LW0
4	77	70.6	265	10	Q39598
5	76	69.7	665	11	Q9QXG2
6	76	69.7	1452	4	Q9H4A0
7	76	69.7	1512	4	Q9H4A1
8	75.5	69.3	531	5	Q24426
9	75.5	69.3	566	5	Q9VF97
10	75	68.8	221	5	Q9VXD3
11	72	66.1	1071	5	Q9W232
12	71	65.1	206	10	Q39597
13	71	65.1	1216	10	Q9SWH3
14	71	65.1	1787	10	Q9M4X9
15	70	64.2	323	4	Q9H782
16	70	64.2	349	4	Q43856

17	70	64.2	423	4	Q75400	075400 homo sapien
18	70	64.2	452	4	Q75404	075404 homo sapien
19	70	64.2	484	11	Q923T4	Q923T4 mus musculus
20	70	64.2	1171	3	Q9P3E2	Q9P3E2 neospora
21	69	63.3	219	11	Q9CS14	Q9CS14 mus musculus
22	69	63.3	1668	5	Q76930	Q76930 drosophila
23	69	63.3	1669	5	Q9V727	Q9V727 drosophila
24	68	62.4	148	5	Q9W4G0	Q9W4G0 drosophila
25	68	62.4	403	16	Q9JW2	Q9JW2 neisseria m
26	68	62.4	545	12	Q91TR2	Q91TR2 tupata herp
27	68	62.4	1340	2	Q9L1H8	Q9L1H8 streptomyces
28	68	62.4	2023	5	Q96542	Q96542 drosophila
29	68	62.4	2023	5	Q9V529	Q9V529 drosophila
30	67.5	61.9	324	5	Q9NG19	Q9NG19 drosophila
31	67.5	61.9	324	5	Q9NG18	Q9NG18 drosophila
32	67.5	61.9	324	5	Q9NG17	Q9NG17 drosophila
33	67.5	61.9	324	5	Q9NG16	Q9NG16 drosophila
34	67.5	61.9	324	5	Q9NG15	Q9NG15 drosophila
35	67.5	61.9	324	5	Q9NG14	Q9NG14 drosophila
36	67	61.5	443	10	Q9C5C6	Q9C5C6 arabidopsis
37	67	61.5	460	5	Q9GZT7	Q9GZT7 caenorhabditis
38	67	61.5	544	3	Q9ETV9	Q9ETV9 neospora
39	67	61.5	544	3	Q9VX54	Q9VX54 drosophila
40	67	61.5	604	5	Q23838	Q23838 drosophila
41	67	61.5	627	10	Q98093	Q98093 arabidopsis
42	67	61.5	671	10	Q9FXA2	Q9FXA2 arabidopsis
43	67	61.5	790	5	Q22872	Q22872 caenorhabditis
44	67	61.5	1037	5	Q9V253	Q9V253 drosophila
45	67	61.5	1203	5	Q21835	Q21835 caenorhabditis

ALIGNMENTS

RESULT ID	1	PRELIMINARY:	PRT:	2451 AA.
Q9VG05	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CG7518 PROTEIN.			
GN	CG7518.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephyroidea; Ephyridiidae; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY;			
RX	MEDLINE-20196006; PubMed-10731132;			
RA	Adams M.D., Ceolnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abri J.F., Aghayani A., An H.-J., Andrews-Finnkooch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Bessley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein F., Brothier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,			
RA	Gosler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,			
RA	Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Jiang Y., Lin X.,
 RA Liu X., Matthei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Mostrelli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003698; AAF54888.2; -
 DR Flybase: FBgn0038108; CG7518.
 DR InterPro: IPR001005; Myb_DNA_bind.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 SQ SEQUENCE 2451 AA; 266959 MW; 086A2293F27481E2 CRC64;

Query Match 77.1%; Score 84; DB 5; Length 2451;
 Best Local Similarity 91.3%; Pred. No. 1.5;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
 DB 1374 AAAAAAAAAAAAAAAAAAAAAA 1396

RESULT 2
 O9EPW8 PRELIMINARY: PRT; 1354 AA.
 AC O9EPW8;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE NISCHARIN.
 GN NISCH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR OUTBRED; TISSUE=BRIN;
 RX MEDLINE=20571837; PubMed=11121431;
 RA Alahari S.K., Lee J.W., Juliano R.L.,
 RT "Nischarin, a Novel Protein That Interacts with the Integrin α 5 β 1
 RT Subunit and Inhibits Cell Migration,"
 RL J. Cell Biol. 151:1141-1154(2000).
 DR EMBL: AF315344; AAG42100.1; -
 DR MGD: MGI:1928323; Nisch.
 DR InterPro: IPR001128; Cyt_P450.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 4.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SMO0370; LRR; 5.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SQ SEQUENCE 1354 AA; 148060 MW; 01BD675FDCA19247 CRC64;

Query Match 73.4%; Score 80; DB 11; Length 1354;
 Best Local Similarity 83.3%; Pred. No. 2.2;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 837 EAPAAAEAPAAAEAPAAAEAPAA 860

RESULT 3
 O91WMO PRELIMINARY: PRT; 110 AA.
 AC O91WMO;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 10.7 KDA PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SALIVARY GLAND;
 RA Strausberg R.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC012681; AAH12681.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 110 AA; 10662 MW; 0581D2F635F87EAB CRC64;

Query Match 72.5%; Score 79; DB 11; Length 110;
 Best Local Similarity 83.3%; Pred. No. 0.32;
 Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
 DB 23 DAAAAAAAAAAAAAAAAAAAAA 46

RESULT 4
 O39598 PRELIMINARY: PRT; 265 AA.
 AC O39598;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CGCR-4. PRODUCT (FRAGMENT).
 GN CGCR-4.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92119224; PubMed=1731966;
 RA Wakarchuk W.W., Muller F.W., Beck C.F.;
 RT "Two GC-rich DNA elements of *Chlamydomonas reinhardtii* with complex
 RT arrangements of directly repeated sequence motifs,"
 RL Plant Mol. Biol. 18:143-146(1992).
 DR EMBL: X17208; CAA35080.1; -
 FT NON_TER
 SQ SEQUENCE 265 AA; 26216 MW; B35318B737CF782 CRC64;

Query Match 70.6%; Score 77; DB 10; Length 265;
 Best Local Similarity 82.6%; Pred. No. 1;
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
 DB 154 AAAAAAAAAAKARVAAEAPAAA 176

RESULT 5
 O9OXG2 PRELIMINARY: PRT; 665 AA.
 AC O9OXG2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CHORODEREMIA PROTEIN.
GN CEM.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;
RT "Cloning and sequencing of the mouse choroaderemia gene."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218084; AAF25478.1; -
DR MGD; MGI:892979; Chm.
DR InterPro: IPR002005; Rab_GDI_REP.
DR Pfam: PF00996; GDI; 1.
DR PRINTS: PRO0891; RABGDIREP.
SQ SEQUENCE 665 AA; 73976 MW; FF71A74AD3FBDE0A CRC64;

Query Match 69.7%; Score 76; DB 11; Length 665;
Best Local Similarity 79.2%; Pred. No. 2.8;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAERAAAAERAAAAA 25
DB 132 EAARAAERAAERAAERAAERAA 155

RESULT 6
Q9H4A0 PRELIMINARY; PRT; 1452 AA.
AC Q9H4A0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDC2L5 PROTEIN KINASE.
GN CDC2L5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RT Picard A., Callebaut I., Perre E., Genevriere A.M.;
RT "A new subfamily of high molecular mass CDC2-related kinases with
RT PITAI/VRP."
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).
DR EMBL; AJ297710; CAC10400.1; -
DR HSSP; P24941; 1BUH.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002290; Ser_thr_PKinase.
DR InterPro: IPR001245; Tyr_PKinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 1452 AA; 158480 MW; C7ED07296B8439CB CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1452;
Best Local Similarity 79.2%; Pred. No. 5.5;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAERAAAAERAAAAA 25
DB 463 EAARAAERAAERAAERAAERAA 486

RESULT 7
Q9H4A1 PRELIMINARY; PRT; 1512 AA.
AC Q9H4A1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDC2L5 PROTEIN KINASE.
GN CDC2L5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RT Picard A., Callebaut I., Perre E., Genevriere A.M.;
RT "A new subfamily of high molecular mass CDC2-related kinases with
RT PITAI/VRP."
RL Biochem. Biophys. Res. Commun. 279:832-837(2001).
DR EMBL; AJ297709; CAC10400.1; -
DR HSSP; P24941; 1BUH.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002290; Ser_thr_PKinase.
DR InterPro: IPR001245; Tyr_PKinase.
DR Pfam: PF00069; Pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 1512 AA; 164969 MW; 283B80553DB57650 CRC64;

Query Match 69.7%; Score 76; DB 4; Length 1512;
Best Local Similarity 79.2%; Pred. No. 5.7;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAERAAAAERAAAAA 25
DB 463 EAARAAERAAERAAERAAERAA 486

RESULT 8
Q24426 PRELIMINARY; PRT; 531 AA.
AC Q24426;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TROPOMYOSIN ISOFORM 33.
GN TMI OR G4898.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84205681; PubMed=6202423;
RA Karik C.C., Mahaffey J.W., Coutu M.D., Fyberg E.A.;
RT "Organization of contractile protein genes within the 88F subdivision
RT of the D. melanogaster third chromosome."
RL Cell 37:469-481(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064486; PubMed=3097506;
RA Karik C.C., Fyberg E.A.;
RT "Two Drosophila melanogaster tropomyosin genes: structural and
RT functional aspects."
RL Mol. Cell. Biol. 6:1965-1973(1986).

DR EMBL: K02621: AAA28968.1: -
DR EMBL: M12840: AAA28968.1: JOINED.
DR EMBL: L00355: AAA28968.1: JOINED.
DR EMBL: L00356: AAA28968.1: JOINED.
DR EMBL: L00357: AAA28968.1: JOINED.
DR EMBL: L00358: AAA28968.1: JOINED.
DR EMBL: L00359: AAA28968.1: JOINED.
DR EMBL: L00360: AAA28968.1: JOINED.
DR EMBL: L00362: AAA28968.1: JOINED.
DR FLYBASE: FBgn0003721: Tml.
DR InterPro: IPR002965: P_rich_extensn.
DR InterPro: IPR000533: Tropomyosin.
DR Pfam: PF00261: Tropomyosin; 2.
DR PRINTS: PR01217: PRICHEXTENS.
DR PRINTS: PR00194: TROPOMYOSIN.
DR PROSITE: PS00326: TROPOMYOSIN; 1.
DR Alternative Splicing.
SQ SEQUENCE 531 AA; 56047 MW; 2619A715E20E4DF CRC64;

Query Match 69.3%; Score 75.5; DB 5; Length 531;
Best Local Similarity 87.5%; Pred. No. 2.6;
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAAAAAAAAAA 25
DB 327 EAAAAEAAAAAE-AEAAAAA 349

RESULT 9
OYVF97 PRELIMINARY; PRT; 566 AA.
ID OYVF97;
AC OYVF97;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TMI PROTEIN.
GN TMI OR CG4898.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Fostler C., Gabor F., Garb N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mount S.M., Moy M., Murphy S., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spleer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster".
RL Science 287:2185-2195(2000).
DR EMBL: AE003708: AAP5164.1: -
DR FLYBASE: FBgn0003721: Tml.
DR InterPro: IPR002965: P_rich_extensn.
DR InterPro: IPR000533: Tropomyosin.
DR Pfam: PF00261: Tropomyosin; 2.
DR PRINTS: PR01217: PRICHEXTENS.
DR PRINTS: PR00194: TROPOMYOSIN.
DR PROSITE: PS00326: TROPOMYOSIN; 1.
SQ SEQUENCE 566 AA; 60449 MW; 2AF07CFB0DA2032E CRC64;

Query Match 69.3%; Score 75.5; DB 5; Length 566;
Best Local Similarity 87.5%; Pred. No. 2.8;
Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAAAAAAAAAA 25
DB 362 EAAAAEAAAAAE-AEAAAAA 384

RESULT 10
OYVXD3 PRELIMINARY; PRT; 221 AA.
ID OYVXD3;
AC OYVXD3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG1012 PROTEIN.
GN CG1012.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Fostler C., Gabor F., Garb N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003503; AAF4641.1; -
DR FlyBase: FBgn030769; CG13012.
SQ SEQUENCE 221 AA; 22987 MW; A1B95919B167C5E2 CRC64;

Query Match 68.8%; Score 75; DB 5; Length 221;
Best Local Similarity 81.8%; Pred. No. 1.4;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 AAAAAAEEEEAAAAA 25
DB 6 AAAAAAEEAATVAEAAVA 27

RESULT 11
Q9W2J2 PRELIMINARY; PRT: 1071 AA.
AC Q9W2J2; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG18375 PROTEIN.
GN CG18375.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agdanyani A., An H.-J., Andrews-Plankko C., Baldwin D.,
RA Ballou R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Bottler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cwley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.K.,
RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003453; AAF4659.1; -
DR HSSP: Q13625; IYCS.
DR FlyBase: FBgn0034606; CG18375.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 2.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00248; ANK; 2.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50088; ANK_REPEAT; 2.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50002; SH3; 1.
KW ANK repeat, Repeat.
SQ SEQUENCE 1071 AA; 115502 MW; F86840E92692B4FA CRC64;

Query Match 66.1%; Score 72; DB 5; Length 1071;
Best Local Similarity 79.2%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAA 25
DB 504 EAAAAAEEEEAAAAAQAEEAA 527

RESULT 12
Q39597 PRELIMINARY; PRT: 206 AA.
AC Q39597; 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CGCR-1 PRODUCT (FRAGMENT).
GN CGCR-1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM15 MT-;
RA MEDLINE=92119224; PubMed=1731966;
RA Wakarchuk W.W., Muller F.W., Beck C.F.,
RT "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
arrangements of directly repeated sequence motifs.";
RL Plant Mol. Biol. 18:143-146(1992).
DR EMBL: X17207; CAA35079.1; -
DR InterPro: IPR001778; POA_allergen.
DR PRINTS: PR00833; POAALLERGEN.
FT NON_TER 1
FT TER 206
SQ SEQUENCE 206 AA; 19869 MW; ED3FF120EF8EFAE1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 206;
Best Local Similarity 75.0%; Pred. No. 3.1;
Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEEEAAAAA 25
DB 49 EAAAAAEEEEAAAAAEEAAPA 72

```

RESULT 13
Q9SMH3 PRELIMINARY; PRT; 1216 AA.
ID Q9SMH3;
AC Q9SMH3;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE VARIABLE FLAGELLAR NUMBER PROTEIN.
GN VFLI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-21GR, CC-1690;
RC MEDLINE=94063476; PubMed=8244002;
RA Tam L.W., Lefebvre P.A.;
RT "Cloning of flagellar genes in Chlamydomonas reinhardtii by DNA
RT insertional mutagenesis."
RL Genetics 135:375-384(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-21GR, CC-1690;
RC MEDLINE=21181843; PubMed=11285274;
RA Sifflow C.D., Lavoie M., Tam L.W., Tousey S., Sanders M., Wu W.C.,
RA Borodovsky M., Lefebvre P.A.;
RT "The vfl1 protein in Chlamydomonas localizes in a rotationally
RT asymmetric pattern at the distal ends of the basal bodies."
RL J. Cell Biol. 153:63-74(2001).
DR EMBL; AF154916; AAD52203.1; -
DR InterPro: IPR004089; Chemotaxis_transducer.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF005560; LRR.
DR SMART: SM00370; LRR.
DR Flagella.
KW SEQUENCE
SQ SEQUENCE 1216 AA; 127943 MW; F06E798B35AF256E CRC64;

Query Match 65.1%; Score 71; DB 10; Length 1216;
Best Local Similarity 73.9%; Pred. No. 14;
Matches 17; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAEEAAAAEEAAAAA 24
Db 635 EAAAGREMAAAVQAAAAELAVA 657

RESULT 14
Q9MAX9 PRELIMINARY; PRT; 1787 AA.
ID Q9MAX9;
AC Q9MAX9;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE FLAGELLAR AUTOTOMY PROTEIN FAIP.
GN FAI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA First R.U., Kim P.J., Griffiths E.R., Quarby L.M.;
RT "Faip is a 171 kDa protein essential for axonemal microtubule severing
RT in Chlamydomonas."
RL J. Cell Sci. 0:0-0(2000).
DR EMBL; AF246990; AAF66419.1; -
DR InterPro: IPR003592; LRR_out.
DR SMART; SM00370; LRR.

```

```

SQ SEQUENCE 1787 AA; 171584 MW; 452A74259E14DC1 CRC64;

Query Match 65.1%; Score 71; DB 10; Length 1787;
Best Local Similarity 81.8%; Pred. No. 20;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AAAAAAEEAAAAEEAAAAA 24
Db 1704 AAAAMEAARAMEAAAALEAAAA 1725

RESULT 15
Q9H782 PRELIMINARY; PRT; 323 AA.
ID Q9H782;
AC Q9H782;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CDNA: FLJ21157 f1s, CLONE CAS09937 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohbayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK024810; BAB15016.1; -
DR InterPro: IPR000104; Antifreeze_1.
DR InterPro: IPR002713; FF.
DR InterPro: IPR001202; WW.
DR Pfam: PF01846; FF.
DR PRINTS: PR00308; ANTIFREEZE1.
DR SMART; SM00441; FF.
DR SMART; SM00456; WW.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
FT NON_TER 323
FT SEQUENCE 323 AA; 35148 MW; EAC7B36489FA074A CRC64;

Query Match 64.2%; Score 70; DB 4; Length 323;
Best Local Similarity 85.7%; Pred. No. 5.6;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 AAAAAAEEAAAAEEAAAAA 25
Db 141 AAEEAAAAVYAAAAA 161

Search completed: July 1, 2002, 06:30:57
Job time: 689 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:22 ; Search time 98.97 Seconds

(without alignments)
28.057 Million cell updates/sec

Title: US-09-461-684-2

Perfect score: 109
Sequence: 1 CEAAAAEAAAAEAAAAEAAAAA 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	109	100.0	25	21	AA13781
2	109	100.0	45	21	AA13783
3	84	77.1	2451	22	ABR71574
4	79	72.5	119	22	ABR40362
5	79	72.5	119	22	AAW61172
6	79	72.5	119	22	AAW73866
7	79	72.5	119	22	AAW34068
8	76	69.7	112	22	ABG15501
9	76	69.7	737	11	AAK05268
10	75.5	69.3	566	22	ABR61040
11	75	68.8	221	22	ABR67690

12	73	67.0	225	21	AAK00758	Human secreted pro
13	72	66.1	54	18	AAW22875	P. americanus skin
14	72	66.1	316	22	ABG18917	Novel human diagno
15	72	66.1	1071	22	ABR60579	Drosophila melanog
16	72	66.1	1444	22	ABG15667	Novel human diagno
17	70	64.2	39	18	AAW22874	P. americanus skin
18	70	64.2	326	22	AAW60765	Gene 16 related pe
19	70	64.2	326	22	AAW60766	Human polypeptide
20	70	64.2	487	22	AAW1402	Human WT1 interact
21	70	64.2	957	22	AAW64150	Synthetic skin typ
22	69	63.3	39	18	AAW22876	P. americanus anti
23	69	63.3	26	18	AAW22881	Novel human diagno
24	69	63.3	1529	17	AAW37985	CORC potassium cha
25	69	63.3	1669	22	ABR54003	Drosophila melanog
26	69	63.3	39	18	AAW22871	P. americanus skin
27	68	62.4	39	18	AAW22869	P. americanus skin
28	68	62.4	148	22	ABR70649	Drosophila melanog
29	68	62.4	1078	22	ABR58620	Drosophila melanog
30	68	62.4	2023	22	ABR63487	Drosophila melanog
31	68	62.4	314	18	AAW14283	Human neuroblastom
32	67	61.5	314	21	AAW70572	Mouse Phox2b prote
33	67	61.5	544	22	ABR61602	Drosophila melanog
34	67	61.5	1037	22	ABR70288	Drosophila melanog
35	67	61.5	1416	22	ABR55499	Drosophila melanog
36	67	61.5	2703	22	ABR63299	Drosophila melanog
37	67	61.5	31	21	AAW8166	Peptide modulating
38	66	60.6	76	17	AAW31273	Elmeria gametocyte
39	66	60.6	76	20	AAW77190	Amino acid fragmen
40	66	60.6	92	21	AAW44712	Shorthorn sculpin
41	66	60.6	372	22	ABR94062	Human protein sequ
42	66	60.6	399	22	ABR3773	Drosophila melanog
43	66	60.6	634	22	ABR71624	Drosophila melanog
44	66	60.6	924	22	ABR67870	Drosophila melanog
45	66	60.6				

ALIGNMENTS

RESULT	1
AA13781	AA13781 standard; peptide: 25 AA.
ID	AA13781
XX	AA13781;
AC	10-NOV-2000 (First entry)
XX	
DT	Soluble peptide antigen pEA.
DE	
XX	peA peptide: cytostatic; vaccine: cytotoxic T cell; CTL; immunotherapy;
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;
KW	prostate; breast; multiple myeloma.
XX	
OS	Unidentified.
XX	
PN	WO200035949-A1.
PD	22-JUN-2000.
XX	
PF	14-DEC-1999; 99WO-US29724.
XX	
PR	14-DEC-1998; 98US-0112324.
PA	(DEND-) DENDREON CORP.
PI	Laus R, Hakim I, Vidovic D;
XX	
PI	WPI: 2000-442365/38.
DR	
XX	
PT	Antigens modified by the covalent addition of a peptide that
PT	facilitates entry into antigen presenting cells, useful for producing
PT	compositions for immunizing against tumors and pathogens -
XX	

PS Claim 2; Page 26; 34pp: English.

XX The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC peptide sequence which facilitate entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is peptide pEA. The modified antigen composition may be used for
CC immunising against, or treating a tumour e.g. prostate and breast
CC carcinoma or multiple myeloma, or pathogen in mammals.

SQ Sequence 25 AA;

Query Match 100.0%; Score 109; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CEAATAAATAAATAAATAAATAA 25
|||||

Db 1 ceataataataataataataataa 25

RESULT 2
AAB13783
ID AAB13783 standard; peptide; 45 AA.
XX
AC AAB13783;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble tandem pEA/ PK peptide conjugate.
XX
KM PK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KM major histocompatibility complex class I; MHC class I; antigen; tumour;
KW prostate; breast; multiple myeloma; pEA peptide.
XX
OS Unidentified.
XX
PN WO200035949-A1.
XX
PD 22-JUN-2000.
XX
PF 14-DEC-1999; 99WO-US29724.
XX
PR 14-DEC-1998; 98US-0112324.
XX
PA (DEND-) DENDREON CORP.
XX
PI Laus R, Hakim I, Vidovic D;
XX
DR WPI: 2000-442365/38.
XX
PT Antigens modified by the covalent addition of a peptide that
PT facilitates entry into antigen presenting cells, useful for producing
PT compositions for immunizing against tumors and pathogens -
XX
PS Claim 2; Page 26; 34pp: English.

XX The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC peptide sequence which facilitate entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is tandem pEA/ PK peptide conjugate. The modified antigen
CC composition may be used for immunising against, or treating a tumour e.g.
CC prostate and breast carcinoma or multiple myeloma, or pathogen in
XX mammals.

SQ Sequence 45 AA;

Query Match 100.0%; Score 109; DB 21; Length 45;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CEAATAAATAAATAAATAAATAA 25
|||||

Db 1 ceataataataataataataataa 25

RESULT 3
ABB71574
ID ABB71574 standard; Protein; 2451 AA.
XX
AC ABB71574;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41514.
XX
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR N-PSDB; ABL15677.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 41514; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (AB01840-ABL16175) and the encoded proteins
CC (AB057737-AB072072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2451 AA;

Query Match 77.1%; Score 84; DB 22; Length 2451;
Best Local Similarity 91.3%; Pred. No. 0.021;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAAAATAAATAAATAAATAA 25
|||||

Db 1374 aataataataataataataataa 1396

```
RESULT 4
ABB40362
ID ABB40362 standard; Peptide; 119 AA.
XX
XX ABB40362;
AC
XX
XX 04-FEB-2002 (first entry)
DT
XX
DE Peptide #7868 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 32997; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 119 AA;
SQ

Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY 2 EAAAAAEEAAAAAEEAAAAA 25
   ||| | ||||| ||||| |||||
DB 31 eaeaeaeaeaeaeaeaeaeae 54

RESULT 5
AAM61172
ID AAM61172 standard; Protein; 119 AA.
XX
XX AAM61172;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 33277.
DE
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW
```

```
XX
XX epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 33277; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 119 AA;
SQ

Query Match 72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY 2 EAAAAAEEAAAAAEEAAAAA 25
   ||| | ||||| ||||| |||||
DB 31 eaeaeaeaeaeaeaeaeaeae 54

RESULT 6
AAM73886
ID AAM73886 standard; Protein; 119 AA.
XX
XX AAM73886;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 34192.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; Leukaemia; Lymphoma; myeloma.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
```

```

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234667.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488930/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
PS Example 4; SEQ ID NO: 34192; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 119 AA;

```

```

Query Match          72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 31 eaeaeaeaeaeaeaeaeaeaeae 54

```

```

RESULT 7
AAM34068
ID AAM34068 standard; Protein: 119 AA.
AC AAM34068;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #8105 encoded by probe for measuring placental gene expression.
XX
KM Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234667.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488937/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

```

```

PS Claim 27; SEQ ID NO 34337; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
SQ Sequence 119 AA;

```

```

Query Match          72.5%; Score 79; DB 22; Length 119;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 EAAAAAAAAAAAAAAAAAAAAA 25
DB 31 eaeaeaeaeaeaeaeaeaeaeae 54

```

```

RESULT 8
ABG15501
ID ABG15501 standard; Protein: 112 AA.
XX
AC ABG15501;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15492.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSO INC.
XX
PI Drmanac RT, Liu C, Tang YR;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS79688.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 45860; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations

```

CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 112 AA;
 SQ

Query Match 69.7%; Score 76; DB 22; Length 112;
 Best Local Similarity 87.0%; Pred. No. 0.0095;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 EAAAAEAAAAEAAAAEAAAAA 25
 ||||| ||||| ||||| |||||
 DB 82 aaaaaaaaaaavaaaaaaaaaa 104

RESULT 9
 ID AAR05268 standard; protein; 737 AA.
 XX AAR05268;
 AC AAR05268;
 XX
 XX 15-AUG-1990 (first entry)
 DT
 XX
 DE Amino acid sequence of human megakaryocytopoietin (MKP)
 DE encoded by upper reading frame of DNA contained in clone 14.
 XX
 KW Human megakaryocytopoietin (MPK); haematopoiesis; immunoassay;
 KW rare leukaemia diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN EP354989-A.
 PN JP03195496-A.
 XX
 PD 21-FEB-1990.
 XX
 PF 27-JUN-1989; 89UP-0111714.
 XX
 PR 28-JUN-1988; 88US-0212623.
 XX
 PA (YISS) YISSUM RES DEV CO.
 XX
 XX Hermona S, Haim Z;
 PI
 XX WPI: 1990-052749/08.
 DR WPI: 1991-282630/40.
 DR N-PDB; AAQ92263.
 XX
 XX New polypeptide hormone megakaryocytopoietin -
 PT involved in haematopoiesis and immune response, and derived
 PT nucleic acid, antibodies, etc., useful eg in diagnosis of leukaemia
 XX
 PS Claim 7, Page 17; Fig 1; 24pp; English.
 XX
 XX A cDNA library constructed from foetal ganglioside mRNA was screened
 CC with a butyryl-choinesterase (bCh) cDNA probe to identify one clone
 CC (clone 14) contg. about 250 bases at the 5'-region of the bCh coding
 CC sequence plus a sequence which hybridised with genomic DNA from patients
 CC with acute myelocytic leukaemia. AAP95268 is the translation of the
 CC upper reading frame and is claimed in the patent. Its nucleic acid can
 CC be isolated, opt. together with its signal sequence, by screening cDNA
 CC or genomic libraries with the clone 14 DNA, to identify a full-length
 CC clone. The material from this clone can be transferred into mammalian or
 CC microbial host cells and these cultured for its prodn. It has hormonal
 CC activity in modulation of haematopoiesis and immune responses. It is
 CC useful in standard immunoassay or hybridisation procedures for
 CC classification and diagnosis of rare leukaemias. It may also be useful
 CC therapeutically.

XX
 SQ Sequence 737 AA;
 XX

Query Match 69.7%; Score 76; DB 11; Length 737;
 Best Local Similarity 79.2%; Pred. No. 0.061;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25
 ||| ||||| ||||| |||||
 DB 106 eaataaakaatkaataaaka 129

RESULT 10
 ID ABB61040 standard; Protein; 566 AA.
 XX ABB61040;
 AC ABB61040;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 9912.
 DE
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PN
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PERE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI: 2001-656860/75.
 DR N-PDB; ABL05143.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 9912; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABR01840-ABR16175) and the encoded proteins
 CC (ABBS7737-ABR72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 566 AA;
 SQ

Query Match 69.3%; Score 75.5; DB 22; Length 566;
 Best Local Similarity 87.5%; Pred. No. 0.054;
 Matches 21; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 2 EAAAAEAAAAEAAAAEAAAAA 25
 ||||| ||||| ||||| |||||
 DB 362 eaataaataaa-eaataaataa 384


```
RESULT 11
AB67690
ID ABB67690 standard; Protein: 221 AA.
XX
XX ABB67690;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 29862.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX
XX (PERE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
PI
XX
XX WPI: 2001-656860/75.
DR
XX
XX N-PSDB; ABL11793.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PR interactions -
XX
XX
XX Disclosure: SEQ ID NO 29862; 21pp + Sequence Listing: English.
PS
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB120511), expressed DNA
CC sequences (AB57737-AB572072).
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
CC
XX
XX Sequence 221 AA;
SQ
Query Match 68.8%; Score 75; DB 22; Length 221;
Best Local Similarity 81.8%; Pred. NO. 0.025;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 AAAAEEAAAEEAAAEEAAA 25
DB 6 aaaaataaataaataaataa 27
RESULT 12
AAG00758
ID AAG00758 standard; Protein: 225 AA.
XX
XX AAG00758;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein, SEQ ID NO: 4839.
DE
XX
XX Human, 5' EST, expressed sequence tag; secreted protein; cDNA isolation;
KW
```

```
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI: 2000-500381/45.
DR
XX
XX N-PSDB; AAC00764.
XX
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PR diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 4839; 71pp + CD-ROM; English.
XX
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
CC
XX
XX Sequence 225 AA;
SQ
Query Match 67.0%; Score 73; DB 21; Length 225;
Best Local Similarity 86.4%; Pred. NO. 0.045;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 AAAAEEAAAEEAAAEEAAA 24
DB 150 aaaaaaaaaaaaaaaaaa 171
RESULT 13
AAW22875
ID AAW22875 standard; Protein: 54 AA.
XX
XX AAW22875;
AC
XX
XX 23-MAR-1998 (first entry)
DT
XX
XX P. americanus skin type antifreeze polypeptide. SAFP8.
DE
XX
XX Antifreeze polypeptide; intracellular skin type AFP; winter flounder;
KW cold resistance; thermal hysteresis; antibacterial.
XX
XX Pseudopleuronectes americanus.
OS
XX
XX WO9728260-A1.
PN
XX
XX 07-AUG-1997.
PD
XX
XX 30-JAN-1997; 97WO-CA00062.
PF
XX
XX 31-JAN-1996; 96US-0010920.
PR
```

XX	(HSCR-) HSC RES & DEV LP.
PA	Gong Z., Hew C;
PI	WPI: 1997-402614/37.
XX	N-Psdb; AAT75505.
DR	
XX	
PT	Isolated fish skin anti-freeze polypeptide - useful to depress
PT	freezing points of aqueous compositions and protect plant, fungal,
PT	animal or bacterial cells from cold
PS	Claim 6; Page 68; 104pp; English.
XS	
CC	This sequence represents the AFP8 clone corresponding to a novel skin
CC	intracellular antifreeze polypeptide (AFP) which induces a concentration
CC	dependent decrease in the freezing point of an aqueous solution. This
CC	novel polypeptide does not contain a signal sequence and is thought to be
CC	intracellular. AFP's can be used to make an aqueous composition resistant
CC	to freezing by changing its thermal hysteresis such as a water or salt
CC	solution, an intracellular compartment of a cell or a food stuff, e.g.
CC	soft serve "frozen" yogurt or ice cream. AFP's can inhibit ice
CC	recrystallisation during cold storage, improving the texture and
CC	palatability of the food and has antibacterial properties. Such
CC	polypeptides can also be expressed to provide cold resistance to cells,
CC	e.g., plant, fungal animal or bacterial cells. The antibodies can be used
CC	to identify and isolate AFP while its promoter can be used to direct
CC	expression of a nucleic acid.
SQ	Sequence 54 AA:
OY	Query Match 66.1%; Score 72; DB 18; Length 54; Best Local Similarity 78.3%; Pred. No. 0.014; Matches 18; Conservative 2; Mismatches 3; Indels 0; Gaps 0.
DB	3 AAAAEEAAAAEFAAAAEFAAAA 25 28 aaaataaaaaakaaatkaaaaa 50
RESULT 14	
ABG18917	
ID	ABG18917 standard; Protein; 316 AA.
AC	ABG18917;
XX	
DT	18-FEB-2002 (first entry)
DE	
XX	Novel human diagnostic protein #18908.
KM	Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
PROS	Homo sapiens.
XX	
PN	WO200175067-A2.
PD	11-OCT-2001.
PF	30-MAR-2001; 2001WO-US08631.
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dimanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
DR	N-Psdb; AASB3104.
PT	New isolated polynucleotide and encoded polypeptides, useful in

```

PT  diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
PS Claim 20; SEQ ID NO 49276; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 316 AA:
XX
OY 3 AAAAAAAAAAAAAAAAAAAAAA 25
   ||||| ||||| ||||| |||||
Db 14 aaaaaaaaaaaaaaaaaaaaaa 36
XX
RESULT 15
ABB60579
ID ABB60579 standard; Protein: 1071 AA.
XX
AC ABB60579;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 8529.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
   pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO2000171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
DR WPI: 2001-656860/75.
XX N-PSDB; ABL04682.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
   genes from Drosophila and for elucidating cell signalling and cell-cell

```

PT Interactions -

XX PS Disclosure; SEQ ID NO 8529; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
 CC sequences (AB57737-AB572072),
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

XX Sequence 1071 AA;

Query Match 66.1%; Score 72; DB 22; Length 1071;

Best Local Similarity 79.2%; Pred. NO. 0.28; Mismatches 5; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAAAAAERAAAAERAAAAERAAAA 25

Db 504 eaaaaaaaaaaaaaaaaaaaaaa 527

Search completed: July 1, 2002, 06:19:23
 Job time: 510 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:10 ; Search time 49.45 Seconds
(without alignments)
46.636 Million cell updates/sec

Title: US-09-461-684-3
Perfect score: 143
Sequence: 1 GGFAGAIAGFIENGEGMIDGMYG 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR71:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	93.7	550	1 HMIIVS2	hemagglutinin prec
2	134	93.7	550	1 HMIIVS3	hemagglutinin prec
3	134	93.7	550	1 HMIIV77	hemagglutinin prec
4	134	93.7	550	1 HMIIV80	hemagglutinin prec
5	134	93.7	550	1 HMIIV33	hemagglutinin prec
6	134	93.7	550	1 HMIIV89	hemagglutinin prec
7	134	93.7	550	1 HMIIV21	hemagglutinin prec
8	134	93.7	550	1 HMIIV98	hemagglutinin prec
9	134	93.7	550	1 HMIIV15	hemagglutinin prec
10	134	93.7	550	2 JQ1153	hemagglutinin prec
11	134	93.7	550	2 JQ1154	hemagglutinin prec
12	134	93.7	550	2 JQ1155	hemagglutinin prec
13	134	93.7	556	1 HMIIVH	hemagglutinin prec
14	134	93.7	556	1 HMIIVHA	hemagglutinin prec
15	134	93.7	556	1 HMIIVHM	hemagglutinin prec
16	134	93.7	556	1 HMIIVDU	hemagglutinin prec
17	133	93.0	561	1 HMIIV49	hemagglutinin prec
18	133	93.0	561	1 HMIIV84	hemagglutinin prec
19	132	92.3	565	1 HMIIVEL	hemagglutinin prec
20	132	92.3	565	1 HMIIVES	hemagglutinin prec
21	132	92.3	566	1 HMIIV6	hemagglutinin prec
22	132	92.3	567	1 HMIIV6	hemagglutinin prec
23	131	91.6	362	2 S38637	hemagglutinin - in
24	131	91.6	550	1 HMIIVT7	hemagglutinin prec
25	131	91.6	550	1 HMIIVT7	hemagglutinin prec
26	131	91.6	565	1 HMIIVT7	hemagglutinin prec
27	131	91.6	565	1 HMIIVT7	hemagglutinin prec
28	131	91.6	565	1 HMIIVT7	hemagglutinin prec
29	131	91.6	565	1 HMIIVT7	hemagglutinin prec

30	131	91.6	565	1 HMIIVT7	hemagglutinin prec
31	131	91.6	565	1 HMIIVT7	hemagglutinin prec
32	131	91.6	565	1 HMIIVT7	hemagglutinin prec
33	131	91.6	565	1 HMIIVT7	hemagglutinin prec
34	131	91.6	565	1 HMIIVT7	hemagglutinin prec
35	131	91.6	565	2 S33703	hemagglutinin - in
36	131	91.6	570	1 A45591	hemagglutinin prec
37	131	91.6	570	2 S22013	hemagglutinin prec
38	131	91.6	570	2 S22014	hemagglutinin prec
39	131	91.6	570	2 S22015	hemagglutinin prec
40	131	91.6	570	2 S22016	hemagglutinin prec
41	131	91.6	570	2 S22017	hemagglutinin prec
42	131	91.6	570	2 S22018	hemagglutinin prec
43	131	91.6	570	2 S22020	hemagglutinin prec
44	131	91.6	570	2 S22021	hemagglutinin prec
45	131	91.6	570	2 S22029	hemagglutinin prec

ALIGNMENTS

RESULT 1

HMIIVS2 hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)

C:Species: Influenza A virus

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998

C:Accession: A29971

R:Kida, H.; Shortridge, K.F.; Webster, R.G.

Virology 162, 160-166, 1988

Article: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China

A:Reference number: A94370; MUID:88101364

A:Accession: A29971

A:Molecule type: genomic RNA

A:Residues: 1-550 <RID>

A:Cross-references: GB:M19056; NID:G324208

A>Note: the sequence in Genbank entry FLAHAB, release 106, (PID:G324209) differs from

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; 110protein; thiolester bond

F:1-328/Product: hemagglutinin HAI #status predicted <HAI>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM>

F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

F:114,466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7% Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGFAGAIAGFIENGEGMIDGMYG 24

DB 330 GGFAGAIAGFIENGEGMIDGMYG 352

RESULT 2

HMIIVS3 hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)

C:Species: Influenza A virus

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998

C:Accession: B29971

R:Kida, H.; Shortridge, K.F.; Webster, R.G.

Virology 162, 160-166, 1988

Article: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China

A:Reference number: A94370; MUID:88101364

A:Accession: B29971

A:Molecule type: genomic RNA

A:Residues: 1-550 <RID>

A:Cross-references: GB:M19057; NID:G324210

A>Note: the sequence in Genbank entry FLAHAB, release 106, (PID:G324211) differs from

C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52,277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;
Best Local Similarity 100.0%: Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGWY 24
DB 330 GLFGAIAGFLENGEGMIDGWY 352

RESULT 3

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A27813
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: A27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16737; NID:g324081; PIDN:AAA43145.1; PID:g324082
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52,277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;
Best Local Similarity 100.0%: Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGWY 24
DB 330 GLFGAIAGFLENGEGMIDGWY 352

RESULT 4

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: B27813
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: B27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16738; NID:g324083
A:Note: The translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52,277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;
Best Local Similarity 100.0%: Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGWY 24
DB 330 GLFGAIAGFLENGEGMIDGWY 352

RESULT 5

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: C27813
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: C27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14,466,52,277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%: Score 134; DB 1; Length 550;
Best Local Similarity 100.0%: Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFLENGEGMIDGWY 24
DB 330 GLFGAIAGFLENGEGMIDGWY 352

RESULT 6

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: D27813
R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: D27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
C:Genetics:
A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:520-536/Domain: transmembrane #status predicted <TM2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1,4e-10; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24
DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

RESULT 7

HMIY21

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998

C:Accession: E27813

R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: E27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16741; NID:9324089

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>

F:7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24
DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

RESULT 8

HMIY98

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998

C:Accession: F27813

R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: F27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16742; NID:9324091

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM1>
F:520-536/Domain: transmembrane #status predicted <TM2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1,4e-10; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24
DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

RESULT 9

HMIY15

hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999

C:Accession: G27813

R:Kida, H.; Kawoka, Y.; Naeye, C.W.; Webster, R.G.

Virology 159, 109-119, 1987

A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.

A:Reference number: A94363; MUID:87265458

A:Accession: G27813

A:Molecule type: genomic RNA

A:Residues: 1-550 <KID>

A:Cross-references: GB:M16743; NID:9324093; PIDN:AAA43149.1; PID:9324094

C:Genetics:

A:Map position: segment 4

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>

F:520-536/Domain: transmembrane #status predicted <TM1>

F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 1,4e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAFGIENGWEGMIDGWTG 24
DB 330 GLFGAIAFGIENGWEGMIDGWTG 352

RESULT 10

JQI153

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: Influenza A virus

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: JQI153

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.

J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian

A:Reference number: JQI153; MUID:91341491

A:Accession: JQI153

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00929; NID:9221279; PIDN:BA00769.1; PID:9221280

A:Note: the authors translated the codon GGG for residue 218 as Gln

A:Note: residues 528-532 are not shown in this publication

C:Superfamily: Influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGAGFIENGWEGMIDGWYG 24
Db 330 GLEFAGAGFIENGWEGMIDGWYG 352

RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/hong kong/10/76) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01154

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491
A:Accession: J01154

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00930; NID:g221273; PIDN:BA00770.1; PID:g221274

A:Note: the authors translated the codon GGG for residue 218 as Glu

C:Keywords: residues 528-532 are not shown in this publication

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-38/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGAGFIENGWEGMIDGWYG 24
Db 330 GLEFAGAGFIENGWEGMIDGWYG 352

RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/hong kong/64/76) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01155

R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991

A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491
A:Accession: J01155

A:Molecule type: genomic RNA

A:Residues: 1-550 <YAS>

A:Cross-references: GB:D00931; NID:g221277; PIDN:BA00771.1; PID:g221278

A:Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538

C:Superfamily: influenza virus hemagglutinin

C:Keywords: glycoprotein; homotrimer

F:1-38/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>

F:8,22,38,165,285,483/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGAGFIENGWEGMIDGWYG 24
Db 330 GLEFAGAGFIENGWEGMIDGWYG 352

RESULT 13

hemagglutinin precursor - influenza A virus

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999

C:Accession: A93705; A93233; A04051; A93231; A94441

R:Both, G.W.; Sleight, M.J.

Nucleic Acids Res. 8, 2561-2575, 1980

A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influ

A:Reference number: A93705; MUID:81053698

A:Accession: A93705

A:Molecule type: genomic RNA

A:Residues: 1-566 <BOT>

A:Cross-references: GB:V01103

A:Experimental source: strain A/NT/60/68/29C

A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A,

R:Dopheide, T.A.; Ward, C.W.

FEBS Lett. 110, 181-183, 1980

A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.

A:Reference number: A91276; MUID:80179105

A:Contents: annotation; disulfide bonds

R:Gethling, M.J.; Bye, J.; Skene, J.; Waterfield, M.

Nature 287, 301-306, 1980

A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes

A:Reference number: A93233; MUID:81030852

A:Accession: A93233

A:Molecule type: genomic RNA

A:Residues: 1-24, 'S', 26, 'D', 28-159, 'G', 161-197, 'I', 199-241, 'L', 243-249 <GET>

A:Experimental source: strain X-31 [H3]

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:556-552/Domain: transmembrane #status predicted <TM1>

F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental

F:555,562,565/Binding site: palmitate (cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEFAGAGFIENGWEGMIDGWYG 24
Db 346 GLEFAGAGFIENGWEGMIDGWYG 368

RESULT 14

hemagglutinin precursor - influenza A virus (strain A/Atch/2/68)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999

C:Accession: A93231; A04051

R:Verhoeven, M.; Pang, R.; Min Jou, W.; Devos, R.; Huybreckx, D.; Saman, E.; Fiers

Nature 286, 771-776, 1980

A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza stra

A:Reference number: A93231; MUID:80254693

A:Accession: A93231

A:Molecule type: genomic RNA

A:Residues: 1-566 <YER>

A:Cross-references: GB:J02090; NID:g324131; PIDN:AAA3178.1; PID:g324132

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
|||||
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

RESULT 15

HM1VHM
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: Influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
C;Accession: A94441; A04051
R;Stoich, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, F
A;Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of
A;Reference number: A94441
A;Accession: A94441
A;Molecule type: genomic RNA
A;Residues: 1-566 <SIDE>
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 93.7%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 1,4e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
|||||
Db 346 GLFGAIGFIENGWEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:11
Job time: 713 sec

33	131	91.6	565	1	HEMA_IAHMI	P15658	influenza a
----	-----	------	-----	---	------------	--------	-------------

Query Match	93.78;	Score 134;	DB 1;	Length 550;
Best Local Similarity	100.08;	Pred. No. 1.1e-10;		

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGEGMIDGWTG 24
|||||
DB 330 GLFGAAGFIENGEGMIDGWTG 352

RESULT 2
HEMA_IADH2 STANDARD; PRT; 550 AA.
ID HEMA_IADH2
AC P12583; Q84011;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11358;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
ducks.";
RT Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M16738; AAA43144.1; -
DR PIR; B27813; HMTV80.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGEGMIDGWTG 24
|||||
DB 330 GLFGAAGFIENGEGMIDGWTG 352

RESULT 3
HEMA_IADH3

ID HEMA_IADH3 STANDARD; PRT; 550 AA.
AC P12584; Q84012; Q89793;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
ducks.";
RT Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M16739; AAA43145.1; -
DR PIR; C27813; HMTV83.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFCB7DE CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAAGFIENGEGMIDGWTG 24
|||||
DB 330 GLFGAAGFIENGEGMIDGWTG 352

RESULT 4
HEMA_IADH4 STANDARD; PRT; 550 AA.
ID HEMA_IADH4
AC P12585; Q84013; Q84014;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/77/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11360;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC -----
DR EMBL: M16740; AAA43146.1; -.
DR PIR: D27813; HMTV89.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2FC8BBD9D0 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGWEGMIDWYG 24
DB 330 GLEGAIGFIENGWEGMIDWYG 352

RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5 STANDARD; PRT; 550 AA.
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11361;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
```

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC -----
DR EMBL: M16741; AAA43147.1; -.
DR PIR: E27813; HMTV21.
DR HSP: F03437; 5HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 368 368 K -> T (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15BF8C CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1,1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAIGFIENGWEGMIDWYG 24
DB 330 GLEGAIGFIENGWEGMIDWYG 352

RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6 STANDARD; PRT; 550 AA.
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
CC Influenza virus A and B group; Influenza A viruses;
CC Influenza A virus.
CC NCBI_TaxID=11362;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M16742; AAA43148.1; -
DR PIR: F27813; HMTV98.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 1
FT CHAIN 1 328
FT CARBOHYD 330 550
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 8 8
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1,1e-10; Length 550;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGEGMIDWYG 24
DQ 330 GLFGAIGFENGEGMIDWYG 352

RESULT 7
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7
AC P12588; Q84018; Q89470;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
CX NCBI_TaxID=11363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
ducks";
RT Virol. 159:109-119(1987).
RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: M16743; AAA43149.1; -
DR PIR: G27813; HMTV15.

InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 1
FT CHAIN 1 328
FT CARBOHYD 330 550
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 8 8
SQ SEQUENCE 550 AA; 61761 MW; 6ER81793281D53EB CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1,1e-10; Length 550;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFENGEGMIDWYG 24
DQ 330 GLFGAIGFENGEGMIDWYG 352

RESULT 8
HEMA_IADHK STANDARD; PRT; 550 AA.
ID HEMA_IADHK
AC P43257;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
CX NCBI_TaxID=11364;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda U., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
of avian H3 influenza viruses to pigs in southern China, where the
A/Hong Kong/68 (H3N2) strain emerged";
RT J. Gen. Virol. 72:2007-2010(1991).
RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: D00929; BAA00769.1; -
DR HSP: P03437; SHM;
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON TER 1 1
FT CHAIN 1 328
FT CARBOHYD 330 550
FT CARBOHYD 22 22
FT CARBOHYD 8 8
FT CARBOHYD 22 22
SQ SEQUENCE 550 AA; 61761 MW; 6ER81793281D53EB CRC64;

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61549 MW; 864639829FE1B89A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAAGFIENGWEGMIDWYG 24
|||||
Db 330 GLEGAAGFIENGWEGMIDWYG 352

RESULT 9
HEMA_IADHL STANDARD; PRT; 550 AA.
AC P43258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=45412;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
of avian H3 influenza viruses to pigs in southern China, where the
A/hong kong/68 (H3N2) strain emerged."
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00931; BA00771.1; -
DR HSSP: P03437; 5HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KM Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61718 MW; A351C56789E4B89A CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLEGAAGFIENGWEGMIDWYG 24
|||||
Db 330 GLEGAAGFIENGWEGMIDWYG 352

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAAGFIENGWEGMIDWYG 24
|||||
Db 330 GLEGAAGFIENGWEGMIDWYG 352

RESULT 10
HEMA_IAGHK STANDARD; PRT; 550 AA.
ID HEMA_IAGHK
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=45414;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
of avian H3 influenza viruses to pigs in southern China, where the
A/hong kong/68 (H3N2) strain emerged."
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00930; BA00770.1; -
DR HSSP: P03437; 5HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KM Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 8 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 22 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61676 MW; 9A1D094DA28BACD2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLEGAAGFIENGWEGMIDWYG 24
|||||
Db 330 GLEGAAGFIENGWEGMIDWYG 352

Query Match 93.7%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
HEMA_IAGH2

ID HEMA_1A2H2 STANDARD: PRT: 550 AA.
 AC P11133: 084019; 084020;
 DT 01-JUL-1988 (Rel. 11, Created)
 DT 15-JUL-1988 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC NCBI_TaxID=11497;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
 RL Virology 162:160-166(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M19057; AAA43212.1; .
 DR PIR: B29971; HMTVS3.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR ProDom: PD000225; Hemagglutn; 1.
 KW Hemagglutinin; Envelope protein; Glycoprotein.
 FT NON_TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
 FT CARBOHYD 330 550
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.le-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGMEGMDGWG 24
 DB 330 GLFGAIGFIENGMEGMDGWG 352

RESULT 12
 HEMA_1A2H3 STANDARD: PRT: 550 AA.
 AC P11134: 084025; 084026;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus;
 OC NCBI_TaxID=11498;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88101364; PubMed=3336940;
 RA Kida H., Shortridge K.F., Webster R.G.;
 RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China."
 RL Virology 162:160-166(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M19056; AAA43211.1; AUL_TERM.
 DR PIR: A29971; HMTVS2.
 DR HSSP: P03437; 2HMG.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR ProDom: PD000225; Hemagglutn; 1.
 KW Hemagglutinin; Envelope protein; Glycoprotein.
 FT NON_TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 330 550
 FT CARBOHYD 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 550 AA; 61580 MW; 991F6DB8C02F24F2 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 550;
 Best Local Similarity 100.0%; Pred. No. 1.le-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGMEGMDGWG 24
 DB 330 GLFGAIGFIENGMEGMDGWG 352

RESULT 13
 HEMA_1A1C STANDARD: PRT: 566 AA.
 AC P03437;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/Alchi/2/68).
 CC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC NCBI_TaxID=150147;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80254693; PubMed=7402351;
 RA Verhoeven M., Fang R., Min Jou W., Devos R., Huylenbroeck D., Saman E., Fiers W.;
 RT "Antigenic drift between the haemagglutinin of the Hong Kong

FT	CHAIN	17	344	HEMAGGLUTININ HA1 CHAIN.
FT	CHAIN	346	566	HEMAGGLUTININ HA2 CHAIN.
FT	CARBOHYD	24	24	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	97	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	566 AA;	63530 MM;	E70F87F0AE1178F4 CRC64;

Query Match 93.7%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
 DB 346 GLFGAIAGFIENGWEGMIDGWYG 368

Search completed: July 1, 2002, 06:31:35
 Job time: 672 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:57 ; Search time 83.98 Seconds
(without alignments)
49.439 Million cell updates/sec

Title: US-09-461-684-3
Perfect score: 143
Sequence: 1 CGLFGAIGFIENGMEGMDGMYG 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:REMBL_19:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:invertebrate:*
7: SP:mammal:*
8: SP:umbc:*
9: SP:organelle:*
10: SP:phage:*
11: SP:plant:*
12: SP:rodent:*
13: SP:virus:*
14: SP:vertebrate:*
15: SP:unclassified:*
16: SP:ivir:*
17: SP:bacteriophage:*
18: SP:archaeo:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	93.7	566	12	Q9DHG0 Influenza a
2	134	93.7	566	12	Q67125 Influenza a
3	134	93.7	566	12	Q67126 Influenza a
4	134	93.7	566	12	Q67132 Influenza a
5	134	93.7	566	12	Q96052 Influenza a
6	134	93.7	566	12	Q91MA7 Influenza a
7	134	93.7	566	12	Q91MA7 Influenza a
8	133	93.0	301	12	Q9DXE3 Influenza a
9	133	92.3	550	12	Q82499 Influenza a
10	132	92.3	550	12	Q82753 Influenza a
11	132	92.3	550	12	Q82498 Influenza a
12	132	92.3	550	12	Q82496 Influenza a
13	132	92.3	571	12	Q03909 Influenza a
14	131	91.6	109	12	Q67050 Influenza a
15	131	91.6	109	12	Q67051 Influenza a
16	131	91.6	109	12	Q67052 Influenza a

17	131	91.6	109	12	Q67053 Influenza a
18	131	91.6	362	12	Q82513 Influenza a
19	131	91.6	362	12	Q82517 Influenza a
20	131	91.6	362	12	Q84174 Influenza a
21	131	91.6	362	12	Q96KD3 Influenza a
22	131	91.6	362	12	Q96KD2 Influenza a
23	131	91.6	362	12	Q96KD1 Influenza a
24	131	91.6	365	12	Q9DL25 Influenza a
25	131	91.6	367	12	Q9DL22 Influenza a
26	131	91.6	368	12	Q9DL29 Influenza a
27	131	91.6	369	12	Q9DL26 Influenza a
28	131	91.6	369	12	Q9DL06 Influenza a
29	131	91.6	369	12	P87689 Influenza a
30	131	91.6	371	12	Q9DL24 Influenza a
31	131	91.6	371	12	P87685 Influenza a
32	131	91.6	373	12	Q9DL20 Influenza a
33	131	91.6	374	12	Q9DL21 Influenza a
34	131	91.6	375	12	Q9DL27 Influenza a
35	131	91.6	375	12	Q9DL05 Influenza a
36	131	91.6	376	12	Q9DL30 Influenza a
37	131	91.6	376	12	Q9DL04 Influenza a
38	131	91.6	377	12	Q9E7P5 Influenza a
39	131	91.6	382	12	Q9DL03 Influenza a
40	131	91.6	408	12	Q9E7P5 Influenza a
41	131	91.6	409	12	Q9E7P5 Influenza a
42	131	91.6	416	12	Q9Q0L4 Influenza a
43	131	91.6	429	12	Q9Q0L4 Influenza a
44	131	91.6	438	12	Q9Q0L3 Influenza a
45	131	91.6	467	12	Q9E312 Influenza a

ALIGNMENTS

RESULT 1
ID Q9DHG0 PRELIMINARY: PRT: 566 AA.
AC Q9DHG0:
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE HAEMAGGLUTININ PRECURSOR.
OS Influenza A virus H3N2.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_taxid=41857;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLONE 7A (H3N2);
RA Mobin M.A., Morris S.J., Smith H., Sweet C.;
RT "Influenza virus-induced apoptosis: a dual role for viral
neuraminidase."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.
CC EMBL: AJ289703; CAC18525.1.
DR HSBP, P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HAEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL: 1
FT POTENTIAL: 16
SQ SEQUENCE 566 AA: 63356 MW: 0BA68192300F72F CRC64;

Query Match 93.7%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDWYG 24
 DB 346 GLFGAIGFIENGEGMIDWYG 368

RESULT 2

ID 067125 PRELIMINARY; PRT; 566 AA.

AC 067125; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

HA. HEMAGGLUTININ.

OS Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3911/92;

RX MEDLINE=95146951; PubMed=7844533;

RA Callan R.J., Early G., Kida H., Hinshaw V.S.;

RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; L31949; AAA64229.1; -.

DR HSSP; P03437; 2V1U.

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTIN12.

DR PRODOM; PD000225; Hemagglutn; 1.

DR Envelope protein; Glycoprotein; Hemagglutinin.

DR SEQUENCE 566 AA; 63456 MW; AE556302A9EB99F CRC64;

SO

Query Match 93.7%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 2.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDWYG 24

DB 346 GLFGAIGFIENGEGMIDWYG 368

RESULT 3

ID 067126 PRELIMINARY; PRT; 566 AA.

AC 067126; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

HA. HEMAGGLUTININ.

OS Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3984/92;

RX MEDLINE=95146951; PubMed=7844533;

RA Callan R.J., Early G., Kida H., Hinshaw V.S.;

RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; L32024; AAA64228.1; -.

DR HSSP; P03437; 2V1U.

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTIN12.

DR PRODOM; PD000225; Hemagglutn; 1.

DR Envelope protein; Glycoprotein; Hemagglutinin.

DR CHAIN 1 346 566 NEURAMINIDASE.

DR SEQUENCE 566 AA; 63441 MW; ESD1B97DE96FECA CRC64;

SO

Query Match 93.7%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 2.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDWYG 24

DB 346 GLFGAIGFIENGEGMIDWYG 368

RESULT 5

ID 098052 PRELIMINARY; PRT; 566 AA.

AC 098052; 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Created)

HA. HEMAGGLUTININ.

OS Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3984/92;

RX MEDLINE=95146951; PubMed=7844533;

RA Callan R.J., Early G., Kida H., Hinshaw V.S.;

RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; L32024; AAA64228.1; -.

DR HSSP; P03437; 2V1U.

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTIN12.

DR PRODOM; PD000225; Hemagglutn; 1.

DR Envelope protein; Glycoprotein; Hemagglutinin.

DR CHAIN 1 346 566 NEURAMINIDASE.

DR SEQUENCE 566 AA; 63441 MW; ESD1B97DE96FECA CRC64;

SO

Query Match 93.7%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 2.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDWYG 24

DB 346 GLFGAIGFIENGEGMIDWYG 368

RESULT 4

ID 067132 PRELIMINARY; PRT; 566 AA.

AC 067132; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

HA. HEMAGGLUTININ.

OS Influenza A virus (strain A/Alchi/2/68).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

CC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=150147;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/ALCHI/2/68;

RX Min J.W., Verhoeven M., Fang R.-X., Devos R., Huybrecock D.,

RA Fiers W.;

RT "Shift and drift in influenza viruses."

RL (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.);

RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,

RL Cambridge University Press, New York (1991).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL; M55059; AAA43239.1; -.

DR HSSP; P03437; 1HGE.

DR InterPro; IPR001364; Hemagglutn.

DR Pfam; PF00509; Hemagglutinin; 1.

DR PRINTS; PR00329; HEMAGGLUTIN12.

DR PRODOM; PD000225; Hemagglutn; 1.

DR Envelope protein; Glycoprotein; Hemagglutinin.

DR CHAIN 1 346 566 NEURAMINIDASE.

DR SEQUENCE 566 AA; 63441 MW; ESD1B97DE96FECA CRC64;

SO

Query Match 93.7%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 2.2e-10;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDWYG 24

DB 346 GLFGAIGFIENGEGMIDWYG 368

```

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=61053698; PubMed=6253883;
RA Both G.W., Sleight M.J.;
RT "Complete nucleotide sequence of the haemagglutinin gene from a human
RT Influenza virus of the Hong Kong subtype.";
RL Nucleic Acids Res. 8:2561-2575(1980).
RN [2]
RP SEQUENCE OF 17-344 FROM N.A.
RX MEDLINE=61194918; PubMed=6164798;
RA Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT subtype: Correlation of amino acid changes with alterations in viral
RT antigenicity.";
RL J. Virol. 37:845-853(1981).
RN [3]
RP SEQUENCE OF 17-566 FROM N.A.
RX MEDLINE=62033276; PubMed=6169843;
RA Both G.W., Sleight M.J.;
RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT Influenza viruses during antigenic drift.";
RL J. Virol. 39:845-853(1981).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: J02135; AAA3189.1; -.
DR HSSP: P03437; 1HGE.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 344 POTENTIAL.
FT CHAIN 346 566 POTENTIAL.
SQ SEQUENCE 566 AA; 63414 MW; C447FD465BE4FCF9 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368

```

RESULT 6
 Q91MA7 PRELIMINARY; PRT; 566 AA.
 AC Q91MA7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ.
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/HONG KONG/1/68(H3N2);
 RX MEDLINE=21287244; PubMed=11371620;
 RT Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;

```

RT "Pattern of mutation in the genome of influenza A virus on adaptation
RT to increased virulence in the mouse lung: Identification of functional
RT themes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
DR EMBL: AF348176; AAK51718.1; -.
SQ SEQUENCE 566 AA; 63387 MW; 01B0D465BE158E1 CRC64;

```

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368

```

RESULT 7
 Q910M5 PRELIMINARY; PRT; 566 AA.
 AC Q910M5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ.
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/HONG KONG/1/68(H3N2);
 RX MEDLINE=21287244; PubMed=11371620;
 RT Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;

Query Match 93.7%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GLFGAIGFIENGWEGMIDGWYG 24
DB 346 GLFGAIGFIENGWEGMIDGWYG 368

```

RESULT 8
 Q9DXE3 PRELIMINARY; PRT; 301 AA.
 AC Q9DXE3;
 DT 01-MAR-2001 (TREMBLrel. 15, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ (FRAGMENT).
 GN HA1.
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=140665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/SHOREBIRD/TAIWAN/31-4/99;
 RA Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: AF11750; AAC33016.1; -
DR InterPro: IPR001364; Hemagglutn.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1
FT CHAIN 1
SQ SEQUENCE 301 AA; 32701 MW; 62A03758B764D57 CRC64;

Query Match 93.0%; Score 133; DB 12; Length 301;
Best Local Similarity 95.7%; Pred. No. 1.5e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWG 24
DB 250 GLFGAIGFIENGWEGMIDGWG 272

RESULT 9
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of Influenza virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule."
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: U08859; AAA16782.1; -
DR HSSP: P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CHAIN 550
SQ SEQUENCE 550 AA; 61772 MW; 50B062B5BFF11FD8 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 4e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWG 24
DB 330 GLFGAIGFIENGWEGMIDGWG 352

RESULT 10
Q82753 PRELIMINARY; PRT; 550 AA.
ID Q82753

AC Q82753:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC unclassified Orthomyxoviridae.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RA MEDLINE-97300854; PubMed-9155874;
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RT virus associated with increased virulence for mice."
RL Arch. Virol. 142:75-88(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RA MEDLINE-97456249; PubMed-9311563;
RA Ward A.C.;
RT "Virulence of Influenza A virus for mouse lung."
RL Virus Genes 14:187-194(1997).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: U08905; AAC79579.1; -
DR HSSP: P03437; 2YIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1
FT CHAIN 1
FT CHAIN 330
FT CHAIN 550
SQ SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 4e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGWG 24
DB 330 GLFGAIGFIENGWEGMIDGWG 352

RESULT 11
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of Influenza virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule."
RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);

RA MEDLINE=88185444; PubMed=3356226;
RT Nakajima S., Takeuchi Y., Nakajima K.;
RT "Location on the evolutionary tree of influenza H3 haemagglutinin
genes of Japanese strains isolated during the 1985-6 season.";
RL Epidemiol. Infect. 100:301-310(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: U08858; AAA18781.1; -.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR Prodom: PD000225; Hemagglutn.1.
KM Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 1
FT CHAIN 1 328 HA1.
FT CHAIN 330 550 HA2.
SQ SEQUENCE 550 AA; 61802 MW; 114413B1CE5A1F6A CRC64;

Query Match 92.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 4e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
1:|||||
DB 330 GIFGAIAGFIENGWEGMIDGMYG 352

RESULT 12
Q82496 PRELIMINARY; PRT; 566 AA.
AC Q82496;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ.
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/SWINE/ANGE-GARDIEN/150/90(H3N2);
RX MEDLINE=95205091; PubMed=7897358;
RA Bikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,
RA Elazhary Y.;
RT "Recent H3N2 swine influenza virus with haemagglutinin and
nucleoprotein genes similar to 1975 human strains.";
RL J. Gen. Virol. 76:697-703(1995).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: U07146; AAA85781.1; -.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR Prodom: PD000225; Hemagglutn.1.
KM Envelope protein; Glycoprotein; Hemagglutinin.
SQ SEQUENCE 566 AA; 63362 MW; 1F88319A567E2FEF CRC64;

Query Match 92.3%; Score 132; DB 12; Length 566;
Best Local Similarity 95.7%; Pred. No. 4.1e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
1:|||||
DB 346 GIFGAIAGFIENGWEGMIDGMYG 368

RESULT 13
Q03909 PRELIMINARY; PRT; 571 AA.
AC Q03909;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ PRECURSOR.
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RA Guo Y., Wang M., Kawacha Y., Gorman O.T., Ito T., Webster R.G.;
RL Submitted (xxx-1991) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
DR EMBL: M65018; AAA43151.1; -.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTIN12.
DR Prodom: PD000225; Hemagglutn.1.
KM Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 349 HA1 CHAIN.
FT CHAIN 350 571 HA2 CHAIN.
FT DISULFID 36 487 BY SIMILARITY.
FT DISULFID 73 298 BY SIMILARITY.
FT DISULFID 85 97 BY SIMILARITY.
FT DISULFID 160 494 BY SIMILARITY.
SQ SEQUENCE 571 AA; 64104 MW; 718DA0F291CE349 CRC64;

Query Match 92.3%; Score 132; DB 12; Length 571;
Best Local Similarity 95.7%; Pred. No. 4.1e-10;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
1:|||||
DB 351 GIFGAIAGFIENGWEGMIDGMYG 373

RESULT 14
Q67050 PRELIMINARY; PRT; 109 AA.
AC Q67050;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/FUKUOKA/C29/85 (H3N2);
RX MEDLINE=81030852; PubMed=7421990;
RA Getling M.-J., Bye J., Skemmel J., Waterfield M.;
RT "Cloning and dna sequence of double-stranded copies of haemagglutinin
genes from h2 and h3 strains elucidates antigenic shift and drift in
human influenza virus.";
RL Nature 287:301-306(1980).
RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-A/FUKUOKA/C29/85 (H3N2);
 RX MEDLINE=93233219; PubMed=7682624;
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
 RT "A common neutralizing epitope conserved between the hemagglutinins of
 influenza A virus H1 and H2 strains";
 RL J. Virol. 67:2552-2558(1993).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: D13581; BAA02776.1; -.
 DR HSSP: P03437; 1H7M.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin.1.
 DR PRINTS: PR00329; HEMAGGLUTN12.
 DR ProDom: PD000225; Hemagglutn.1.
 KW Envelope protein; glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;
 Best Local Similarity 91.3%; Pred. No. 8.9e-11;
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMDGNYG 24
 DB 40 GIFGAIAGFIENGWEGMDGNYG 62

RESULT 15
 ID O67051 PRELIMINARY; PRT; 109 AA.
 AC O67051;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HEMAGGLUTININ (FRAGMENT).
 GN HA.
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=11320;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-A/SICHUAN/2/87 (H3N2);
 RX MEDLINE=81030852; PubMed=7421990;
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin
 genes from h2 and h3 strains elucidates antigenic shift and drift in
 human influenza virus";
 RL Nature 287:301-306(1980).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN-A/SICHUAN/2/87 (H3N2);
 RX MEDLINE=93233219; PubMed=7682624;
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
 RT "A common neutralizing epitope conserved between the hemagglutinins of
 influenza A virus H1 and H2 strains";
 RL J. Virol. 67:2552-2558(1993).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: D13582; BAA02777.1; -.
 DR HSSP: P03437; 1HGE.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin.1.
 DR PRINTS: PR00329; HEMAGGLUTN12.

DR ProDom: PD000225; Hemagglutn.1.
 KW Envelope protein; glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 91.6%; Score 131; DB 12; Length 109;
 Best Local Similarity 91.3%; Pred. No. 8.9e-11;
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIAGFIENGWEGMDGNYG 24
 DB 40 GIFGAIAGFIENGWEGMDGNYG 62

Search completed: July 1, 2002, 06:30:58
 Job time: 690 sec

Mon Jul 1 13:54:59 2002

us-09-461-684-3.rpt

Page 7

RESULT	1
ID	AA13782 standard; peptide; 24 AA.
XX	AA13782
AC	AA13782;
XX	10-NOV-2000 (first entry)
DT	
XX	
DE	Soluble peptide antigen HA.
XX	
XX	HA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW	major histocompatibility complex class 1; MHC class 1; antigen; tumour
KM	prostate; breast; multiple myeloma.
XX	
XX	Unidentified.
OS	
XX	W0200035949-A1.
FN	
XX	22-JUN-2000.
PD	
XX	14-DEC-1999; 99WO-US29724.
PF	
XX	14-DEC-1998; 98US-0112324.
PR	
XX	
PA	(DEND-) DENDREON CORP.
XX	
XX	Laus R, Hakim I, Vidovic D;
PI	
XX	WPI; 2000-442365/38.
DR	
XX	
XX	
PT	Antigens modified by the covalent addition of a peptide that
PT	facilitates entry, into antigen presenting cells, useful for producing
PT	compositions for immunizing against tumors and pathogens -
XX	

PS Claim 2; Page 26; 34pp; English.

XX The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC response. The protein antigen is modified by the covalent addition of a
CC peptide sequence which facilitate entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is peptide HA. The modified antigen composition may be used for
CC immunising against, or treating a tumour e.g. prostate and breast
CC carcinoma or multiple myeloma, or pathogen in mammals.

XX
SQ Sequence 24 AA;

Query Match 100.0%; Score 143; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.2e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIFENGMEGMDGWG 24
1 cglfgaifengmegmdgwyg 24

RESULT 2

AAW34269 ID AAW34269 standard; peptide; 26 AA.

AC AAW34269;

DT 26-MAY-1998 (first entry)

DE Influenza fusion peptide #1.

KW Haemagglutinin; cystein; liposome.

OS Synthetic.

PN EPA97997-A.

PD 12-AUG-1992.

PF 02-FEB-1991; 91EP-0101414.

PR 02-FEB-1991; 91EP-0101414.

PA (NIKA-) NIKA HEALTH PROD LTD.

PI Gluck R, Herrmann P, Klein P;

WI: 1992-270078/33.

PT Dryer-conc. phospholipid bilayer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS

XX Disclosure; Fig 2; 13pp; English.

XX The sequences given in AAW34269-19 are influenza fusion peptides,
CC derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cystein residues at one
CC end of these peptides has been found useful for the fusion activity,
CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.

XX Sequence 26 AA;

Query Match 100.0%; Score 143; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIFENGMEGMDGWG 24
3 cglfgaifengmegmdgwyg 26

RESULT 3

AAW34269 ID AAW34269 standard; peptide; 26 AA.

AC AAW34269;

DT 14-MAY-1998 (first entry)

DE Synthetic lipid vesicle fusion peptide 1.

KW Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;

KW drug delivery system; membrane; gene therapy; diagnosis; treatment;

KW cancer; leukaemia; viral infection.

OS Synthetic.

PN WO9741834-A1.

PD 13-NOV-1997.

PF 04-MAY-1997; 97WO-EP02268.

PR 08-MAY-1996; 96EP-0107282.

PA (NIKA-) NIKA HEALTH PROD LTD.

PI Glueck R, Klein P, Maeltl ER;

WI: 1997-558673/51.

PT Vesicle with cationic lipid bilayer that includes viral fusion
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukaemia and viral infections

PS Disclosure; Page 10; 52pp; English.

XX Peptides AAW34269-W34284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment)
CC of cancer, leukaemia and viral infections in humans or animals. Genetic
CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.

XX Sequence 26 AA;

Query Match 100.0%; Score 143; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24
Db 3 cglfgaigfiengwewgmldgwyg 26

RESULT 4
AAB70090
ID AAB70090 standard; peptide: 26 AA.
XX
XX AAB70090;
DT 14-MAY-2001 (first entry)
DE Cationic virosome crosslinker #1 for polypeptide binding.
XX
XX Cationic virosome; cytosolic; gene therapy; lipid bilayer vesicle;
KW viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;
KM crosslinker.
XX
XX Unidentified.
XX
XX NZ504444-A.
XX
XX 24-NOV-2000.
XX
XX 10-MAY-2000; 2000NZ-0504444.
XX
XX 10-MAY-2000; 2000NZ-0332666.
XX
XX (NIKA-) NIKA HEALTH PROD LTD.
XX
XX Walti ER, Gluck R, Klein P;
DR WPI; 2001-233042/24.
XX
XX Lipid bilayer vesicle, useful for delivering drugs to target cells such
PT as cancer, leukemic, or virally infected cells, comprises viral
PT glycoproteins in positively charged membrane
XX
XX PS Disclosure; Page 14; 41pp; English.

CC The present sequence is a crosslinker which is capable of linking to the
CC surface of a novel cationic virosome and is capable of binding
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising
CC a membrane with a net positive charge and contains 5-30 weight % based on
CC total lipid, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amine
CC (DOPSPR), together with other lipids and at least one active fusogenic
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the
CC vesicle to be internalised by target cells through phagocytosis or
CC endocytosis. The virosome is useful for delivering a desired drug or
CC substance, preferably a nucleic acid, to target cells (resting or
CC proliferating mammalian cells) such as cancer cells, leukemic cells or
CC virally infected cells in vitro, in diagnostic or medical applications
CC and for the manufacture of medicament for carrying out non-infectious
CC gene therapy.
XX
XX SQ Sequence 26 AA;

Query Match 100.0%; Score 143; DB 22; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.4e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24
Db 3 cglfgaigfiengwewgmldgwyg 26

RESULT 5
AAB13784
ID AAB13784 standard; peptide: 44 AA.
XX
XX AAB13784;

XX 10-NOV-2000 (first entry)
DT
XX Soluble tandem HA/ PK peptide conjugate.
DE
XX PK peptide; cytotoxic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW major histocompatibility complex class I; MHC class I; antigen; tumour;
KW prostate; breast; multiple myeloma; HA peptide.
XX
XX Unidentified.
XX
XX WO200035949-A1.
XX
XX 22-JUN-2000.
XX
XX 14-DEC-1999; 99WO-US29724.
XX
XX 14-DEC-1998; 98US-0112324.
XX
XX (DEND-) DENDREON CORP.
XX
XX Laus R, Hakim I, Vidovic D;
XX
XX WPI; 2000-442365/38.
XX
XX Antigen modified by the covalent addition of a peptide that
PT facilitates entry into antigen presenting cells, useful for producing
PT compositions for immunizing against tumors and pathogens
XX
XX PS Claim 2; Page 26; 34pp; English.

CC The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC response. The protein antigen is modified by the covalent addition of a
CC peptide sequence which facilitate entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is tandem HA/ PK peptide conjugate. The modified antigen
CC composition may be used for immunising against, or treating a tumour e.g.
CC prostate and breast carcinoma or multiple myeloma, or pathogen in
CC mammals.
XX
XX SQ Sequence 44 AA;

Query Match 100.0%; Score 143; DB 21; Length 44;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGLFGAIGFIENGWEGMIDGWY 24
Db 1 cglfgaigfiengwewgmldgwyg 24

RESULT 6
AAM29226
ID AAM29226 standard; peptide: 23 AA.
XX
XX AAM29226;
XX
XX 02-MAR-1998 (first entry)
XX
XX Membrane active acidic peptide INF6 from WO9730170.
DE Membrane active; acidic peptide; cationic lipid; transfection;
KW gene therapy; tumour; vaccine; cytokine gene.
XX
XX Synthetic.
XX
XX Influenza virus.
XX
XX PN WO9730170-A1.

PD 21-AUG-1997.
XX
XX 13-FEB-1997; 97WO-EP00649.
PE
XX 15-FEB-1996; 96DE-4005548.
PR
XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.
PA
XX Kichler A, Mechtler K, Wagner E;
PI
XX WPI; 1997-425044/39.
DR
XX
XX Composition containing nucleic acid complexed with cationic lipid
PT and a membrane-active, acidic peptide - used to increase efficiency
PT of transfection by destabilising the endosomal membrane, especially
PT useful for gene therapy
XX
XX
PS Claim 4; Page 33; 63pp; German.
XX
XX A new composition has been developed for transfecting higher eukaryotic
CC cells contains: (i) a complex of nucleic acid to be expressed plus a
CC suboptimal concentration of one or more cationic lipids, and optionally
CC at least one auxiliary lipid; and (ii) the new feature, at least one
CC membrane-active, acidic peptide; where the ratio total positive
CC charges:total-negative charges in the composition is about 0-3,
CC preferably 0-2. The present sequence represents a specifically claimed
CC transfection for use in the composition. The composition is used for
CC transfection in vitro, in vivo or ex vivo, particularly for gene
CC therapy. A typical application is production of tumor vaccines from
CC autologous cells transfected with a cytokine gene. The acidic peptide
CC destabilises the endosomal membrane and significantly increases
CC transfection efficiency, e.g. by as much as 1000 times, equivalent to
CC that achieved with the optimal amount of cationic lipid. The acidic
CC peptide also makes the transfection complex less sensitive to serum.
XX
XX Sequence 23 AA:
SQ
Query Match 93.7%; Score 134; DB 18; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.6e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLEGAIAGFIEGMEGMDWYG 24
Db 1 glfgaiafienegwgmldwyg 23
RESULT 7
AAR98016
ID AAR98016 standard; peptide: 25 AA.
XX
XX AAR98016;
AC
XX
XX 12-NOV-1996 (first entry)
DT
XX
XX Fusedogenic peptide derived from HA-2 spike glycoprotein.
DE
XX
XX Fusedogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMPC;
KM endosome membrane disruption promoting component; cationic polyamine;
KM multifunctional molecular complex; foodstuff; herbicide; insecticide;
KW plant growth regulator; miticide; rodenticide; fungicide; parasiticide;
KW nematocide; immunisation; pathogen; therapy; autoimmune disease;
KW hyperproliferating disease.
XX
XX Synthetic.
OS
XX
XX WO9610038-A1.
PN
XX
XX 04-APR-1996.
PD
XX
XX 28-SEP-1995; 95WO-US12502.
PF
XX
XX 28-SEP-1994; 94US-0314060.
PR

XX
XX (APOL-) APOLLON INC.
PA
XX
XX Boutin RH;
PI
XX
XX WPI; 1996-200887/20.
DR
XX
XX
XX New complexes for nucleic acid transfer to target cells - comprising
PT a nucleic acid compn. and a cationic polyamine with an endosome
PT membrane disruption component.
XX
XX
PS Disclosure; Page 30; 138pp; English.
XX
XX AAR98010-R98041 represent fusedogenic peptides derived from spike
CC glycoproteins of enveloped viruses. These fusedogenic peptides act as
CC endosome membrane disruption promoting components (EMDPs). These
CC sequences were used in a transfer moiety which makes up part of the
CC multifunctional molecular complex of the invention. The multifunctional
CC molecular complex is used for the transfer of a nucleic acid composition
CC to a target cell and comprises the nucleic acid composition and a
CC transfer moiety comprising one or more cationic polyamine components
CC bound to the nucleic acid composition. There are several different
CC cationic polyamines that can be used in the complex, one of which
CC contains one of these sequences linked via a bridging group to the main
CC body of the cationic polyamine. The products can be used for delivering
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth
CC regulators, insecticides, miticides, rodenticides, fungicides,
CC parasiticides or nematocides. They can also be used for immunising an
CC individual against a pathogen or for treating an autoimmune disease or
CC hyperproliferating disease. The complexes provide for a high level of
CC transfection and expression of the nucleic acid molecules in target
CC cells.
XX
XX Sequence 25 AA:
SQ
Query Match 93.7%; Score 134; DB 17; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.6e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLEGAIAGFIEGMEGMDWYG 24
Db 1 glfgaiafienegwgmldwyg 23
RESULT 8
AAR26005
ID AAR26005 standard; peptide: 26 AA.
XX
XX AAR26005;
AC
XX
XX 26-JAN-1993 (first entry)
DT
XX
XX Influenza fusion peptide #2.
DE
XX
XX Haemagglutinin; cysteine; liposome.
KM
XX
XX Synthetic.
OS
XX
XX BP49797-A.
PN
XX
XX 12-AUG-1992.
PD
XX
XX 02-FEB-1991; 91EP-0101414.
PF
XX
XX 02-FEB-1991; 91EP-0101414.
PR
XX
XX (NIKA-) NIKA HEALTH PROD LTD.
PA
XX
XX Gluck R, Herrmann P, Klein P;
PI
XX
XX WPI; 1992-270078/33.
DR
XX

PT Drug-contg. phospholipid bilayer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS
PS Disclosure; Fig 2; 13pp; English.
XX The sequences given in AAR26004-19 are influenza fusion peptides,
CC derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cysteine residues at one
CC end of these peptides has been found useful for the fusion activity,
CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.
XX
SQ Sequence 26 AA;
Query Match 93.7%; Score 134; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAAGFIENGWEGMIDGNYG 24
Db 1 glfgaagfiengwegmidgwyg 23
RESULT 9
AAM34270
ID AAM34270 standard; peptide; 26 AA.
XX
AC AAM34270;
XX
DT 14-MAY-1998 (first entry)
XX
DE Synthetic lipid vesicle fusion peptide 2.
XX
XX Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;
KW drug delivery system; membrane; gene therapy; diagnosis; treatment;
KW cancer; leukaemia; viral infection.
XX
OS Synthetic.
XX
PN W09741834-A1.
XX
PD 13-NOV-1997.
XX
PF 04-MAY-1997; 97WO-EP02268.
XX
PR 08-MAY-1996; 96EP-0107282.
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
PI Glueck R, Klein P, Waelti ER;
XX
DR WPI; 1997-558673/51.
XX
XX Vesicle with cationic lipid bilayer that includes viral fusion
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukaemia and viral infections
XX
PS Disclosure; Page 10; 52pp; English.
XX
CC Peptides AAM4269-W34284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment),
CC of cancer, leukaemia and viral infections in humans or animals. Genetic

CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.
XX
SQ Sequence 26 AA;
Query Match 93.7%; Score 134; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAAGFIENGWEGMIDGNYG 24
Db 1 glfgaagfiengwegmidgwyg 23
RESULT 10
AAB70091
ID AAB70091 standard; peptide; 26 AA.
XX
AC AAB70091;
XX
DT 14-MAY-2001 (first entry)
XX
DE Cationic virosome crosslinker #2 for polypeptide binding.
XX
XX Cationic virosome; cytostatic; gene therapy; lipid bilayer vesicle;
KW viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;
KW crosslinker.
XX
XX Unidentified.
XX
PN N250444-A.
XX
PD 24-NOV-2000.
XX
PF 10-MAY-2000; 2000NZ-0504444.
XX
PR 10-MAY-2000; 2000NZ-0332666.
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
PI Walti ER, Glueck R, Klein P;
XX
DR WPI; 2001-233042/24.
XX
XX Lipid bilayer vesicle, useful for delivering drugs to target cells such
PT as cancer, leukemic, or virally infected cells, comprises viral
PT glycoproteins in positively charged membrane -
XX
PS Disclosure; Page 14; 41pp; English.
XX
XX The present sequence is a crosslinker which is capable of linking to the
CC surface of a novel cationic virosome and is capable of binding
CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising
CC a membrane with a net positive charge and contains 5-30 weight % based on
CC total lipids, 1,3-dioleoyloxy-2-(6-carboxy-spermyl)-propyl-amide
CC (DOSER), together with other lipids and at least one active fusogenic
CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the
CC vesicle to be internalised by target cells through phagocytosis or
CC endocytosis. The virosome is useful for delivering a desired drug or
CC substance, preferably a nucleic acid, to target cells (resting or
CC proliferating mammalian cells) such as cancer cells, leukemic cells or
CC virally infected cells in vitro, in diagnostic or medical applications

CC and for the manufacture of medicament for carrying out non-infectious
CC gene therapy.

CC Sequence 26 AA;

Query Match

Best Local Similarity 93.7%; Score 134; DB 22; Length 26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWY 24

Db 1 glfgaigfiengwegmidgwyg 23

RESULT 11

AAR98024 ID AAR98024 standard; peptide; 28 AA.

AC AAR98024;

DT 12-NOV-1996 (first entry)

DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.

KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMPDC;

KW multimeric membrane disruption promoting component; cationic polyanime;

KW plant growth regulator; multicide; rodenticide; fungicide; parasiticide;

KW nematocide; immunisation; pathogen; therapy; autoimmune disease;

KW hyperproliferating disease.

OS Synthetic.

PN WO9610038-A1.

PD 04-APR-1996.

PF 28-SEP-1995; 95WO-US12502.

PR 28-SEP-1994; 94US-0314060.

PA (APOL-) APOLLON INC.

PI Boutin RH;

DR WPI; 1996-200887/20.

PT New complexes for nucleic acid transfer to target cells - comprising

PT membrane disruption component.

PS Disclosure; Page 30; 138pp; English.

CC AAR98010-988041 represent fusogenic peptides derived from spike

CC glycoproteins of enveloped viruses. These fusogenic peptides act as

CC endosome membrane disruption promoting components (EMPDCs). These

CC sequences were used in a transfer moiety which makes up part of the

CC multifunctional molecular complex of the invention. The multifunctional

CC molecular complex is used for the transfer of a nucleic acid composition

CC to a target cell, and comprises the nucleic acid composition and a

CC transfer moiety comprising one or more cationic polyanime components

CC bound to the nucleic acid composition. There are several different

CC cationic polyanimes that can be used in the complex, one of which

CC contains one of these sequences linked via a bridging group to the main

CC body of the cationic polyanime. The products can be used for delivering

CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth

CC regulators, insecticides, miticides, rodenticides, parasiticides,

CC parasiticides or nematocides. They can also be used for immunising an

CC individual against a pathogen or for treating an autoimmune disease or

CC hyperproliferating disease. The complexes provide for a high level of

CC transfection and expression of the nucleic acid molecules in target

CC cells.

XX Sequence 28 AA;

Query Match

Best Local Similarity 93.7%; Score 134; DB 17; Length 28;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWY 24

Db 1 glfgaigfiengwegmidgwyg 23

RESULT 12

AAR98680 ID AAR98680 standard; peptide; 30 AA.

AC AAR98680;

DT 23-MAY-2000 (first entry)

DE Core polypeptide fragment T No. 1242.

KW Retrovirus; hybrid polypeptide; enhancer; gp41; envelope protein; HIV-1;

KW HIV-2; HIV; pharmacokinetic; half-life; growth factor; cytokine; viral;

KW anti-fusogenic; differentiation factor; interleukin; interferon;

KW colony stimulating factor; hormone; angiogenic factor.

OS Unidentified.

PN WO9959615-A1.

PD 25-NOV-1999.

PF 20-MAY-1999; 99WO-US11219.

PR 20-MAY-1998; 98US-0082279.

PA (TRIM-) TRIMERIS INC.

PI Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;

DR WPI; 2000-136792/12.

PT A new hybrid polypeptide with enhanced pharmacokinetic properties

PT comprises enhancer sequence -

PS Disclosure; Page 42; 124pp; English.

CC The invention relates to hybrid polypeptides comprising enhancer peptide

CC sequence linked to core polypeptides. The enhancer polypeptides are

CC derived from various retroviral envelope (gp41) protein sequences,

CC especially from HIV-1, HIV-2 and HIV. The enhancer peptides enhance the

CC pharmacokinetic properties such as increasing the half-life of any core

CC polypeptide that they are linked to. The core polypeptides are of any

CC function that may be introduced into a living system and that can

CC function as a pharmacologically useful peptide for the treatment or

CC prevention of a disease. The core polypeptides are bioactive peptides

CC selected from a growth factor, cytokine, differentiation factor,

CC interleukin, interferon, colony stimulating factor, hormone or

CC angiogenic factor. The peptides of the invention can be used for

CC inhibiting viral infection and can be used in anti-viral and

CC anti-fusogenic treatments. Sequences AAR98651-990055 represent core

CC polypeptide fragments that can be used in the invention. Some sequences

CC among those indicated also comprise enhancer fragments at terminal ends

CC and form hybrid polypeptides.

CC Sequence 30 AA;

Query Match

Best Local Similarity 93.7%; Score 134; DB 21; Length 30;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGEGMIDGWTG 24
|||||
Db 1 glfgaigfiengwegmidgwyg 23

RESULT 13

AB01088
ID AB01088 standard; Peptide: 30 AA.

XX ABB01088;

DT 03-JAN-2002 (first entry)

DE Viral DP178/107-like region peptide T1242.

XX Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;
KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;
infection.

XX Viridiae.

OS Viridiae.

FT Modified-site 1 Location/Qualifiers

FT Modified-site 30 /note="N-terminal is substituted by Ac"

FT Modified-site 30 /note="C-terminal amide"

PN WO200164013-A2.

XX 07-SEP-2001.

PD 07-FEB-2001; 2001WO-US03988.

PE 29-FEB-2000; 2000US-0515965.

PR (TRIM-) TRIMERIS INC.

XX Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;

PI WPI; 2001-514829/56.

DR Heptad repeat region peptide analogs useful for inhibiting virus/cells

PT fusion, useful for treating HIV and Respiratory Syncytial virus

PT Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

AB02537
ID AB02537 standard; Peptide: 30 AA.

XX ABB02537;

DT 03-JAN-2002 (first entry)

DE Viral core polypeptide, SEQ ID NO: 1064.

XX Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;
KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;
infection.

XX Viridiae.

OS WO200164013-A2.

PN 07-SEP-2001.

PD 07-FEB-2001; 2001WO-US03988.

PE 29-FEB-2000; 2000US-0515965.

PR (TRIM-) TRIMERIS INC.

XX Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;

PI WPI; 2001-514829/56.

DR Heptad repeat region peptide analogs useful for inhibiting virus/cells

PT fusion, useful for treating HIV and Respiratory Syncytial virus

PT Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

AB02537
ID AB02537 standard; Peptide: 30 AA.

XX ABB02537;

DT 03-JAN-2002 (first entry)

DE Viral core polypeptide, SEQ ID NO: 1064.

XX Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;
KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;
infection.

XX Viridiae.

OS WO200164013-A2.

PN 07-SEP-2001.

PD 07-FEB-2001; 2001WO-US03988.

PE 29-FEB-2000; 2000US-0515965.

PR (TRIM-) TRIMERIS INC.

XX Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;

PI WPI; 2001-514829/56.

DR Heptad repeat region peptide analogs useful for inhibiting virus/cells

PT fusion, useful for treating HIV and Respiratory Syncytial virus

PT Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

XX Infection -

OS Synthetic.

FH	Key	Location/Qualifiers

/note= "N-terminal is substituted by Ac"

```

FT      /note= "C-terminal amide"
  
```

PN WO200151673-A2.

PD 19-JUL-2001.
yy

PE 05-JUL-2000;
YYPR 09-JUL-1999; 99US-0350841.
XX

PA (JRLM-) JRLMERIS INC.
XX

XX
P1 JELLS P, LACKEY JW, ELICKSON JB, LAWLESS MK, METULKA G,
XX

WFL; 2001-442131/41-
DK XX

PT	DPI107/DPI178 complex,
F1	identifying a component

PT DP107/DP178 complex -

PS Disclosure; Page 74; 259pp; English.

CC The present invention relates to peptides which exhibit anti-retroviral

CC DP178-like and DP107-like peptides. The DP178 peptide corresponds

CC immunodeficiency virus 1 (HIV-1) isolate LAI. The DP107 peptide

CC also relates to a method of identifying compounds that inhibit the

CC detecting the formation of a DP107/DP178 complex, both in the presence

CC and DPI78 peptides. The method is useful for identifying compounds,

CC antitumor or intracellular modulatory activity. The

CC retroviral, particularly HIV, transmission to uninfected cells. The

CC of the invention.
yy

sq Sequence 30 AA;

Query Match . Q3 79: Score 134: PB 23: Length 30:

Best Local Similarity	100.0%;	Pred. No. 3.6e-13;			
Matches	23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

0Y 2 GLEGATAGETENGWEGMTDGYG 24

Db 1 afafienawamidawa 23

Search completed: July 1, 2002, 06:19:24

0
200
C
TTC
CCE

Query Match	93.78;	Score 134;	DB 22;	Length 30;
Best Local Similarity	100.00;	Prod No	3 60-13;	

Best Local Similarity 100.0%; Pred. NO. 3.6e-13;
Matches 23; Conservative 0; Mismatches 0; Indels 0;
Gaps 0

2 Y
2 GLFGAIAGFIENGWEGMIDGWY 24

Db 1 glfgalagfiengwegmidgwyg 23

Search completed: July 1, 2002, 06:19:24
Job time: 511 sec

A:Gene: ATSP:T20N10.250
 A:Map position: 3
 A:Intons: 312/3; 359/3; 444/3
 C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

Query Match 48.8%; Score 102; DB 2; Length 517;
 Best Local Similarity 80.8%; Pred. No. 0.17; 5; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 20 EAAAAA...
 Db 440 ERVGKKKKKKKKKKKKKKKKKKKK 465

RESULT 3
 T39683
 Zucotin-like protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
 C:Accession: T39683; T40195
 R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 Submitted to the EMBL Data Library, March 1998
 A:Reference number: Z21869
 A:Accession: T39683
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-124 <OL1>
 A:Cross-references: EMBL:AL049489; PIDN:CA539796.1; GSPDB:GN00067; SPDB:SPBC1778.01c
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
 Submitted to the EMBL Data Library, February 1998
 A:Reference number: Z21910
 A:Accession: T40195
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 89-442 <WOO>
 A:Cross-references: EMBL:Z97992; PIDN:CA810796.1; GSPDB:GN00067; SPDB:SPBC30D10.01
 C:Experimental source: strain 972h-1; cosmid c30D10
 C:Genetics:
 A:Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01
 A:Map position: 2

Query Match 48.8%; Score 101.5; DB 2; Length 442;
 Best Local Similarity 56.0%; Pred. No. 0.17;
 Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;

QY 2 EAAAAA...
 Db 309 EAAAAA... 358

RESULT 4
 I52523
 Nucleoporin p62 homolog - rat (fragment)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
 C:Accession: I52523
 R:Wang, Z.-Q.; Akmal, K.M.; Kim, K.H.
 Biol. Reprod. 51, 1022-1030, 1994
 A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the ger
 A:Reference number: I52523; MUID:95151924
 A:Accession: I52523
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-215 <RBS>
 A:Cross-references: GB:S75397; NID:913245; PIDN:AA33384.1; PID:913246
 A:Experimental source: testis

Query Match 47.8%; Score 100; DB 2; Length 215;
 Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 KKKKKKKKKKKKKKKKKKKKK 45
 Db 35 KKKKKKKKKKKKKKKKKKKKK 54

RESULT 5
 T46395
 Hypothetical protein DKFZp434i1120.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46395
 R:Ottenwälder, P.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wilmann, S.
 Submitted to the Protein Sequence Database, January 2000
 A:Reference number: Z23031
 A:Accession: T46395
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-380 <AAA>
 A:Cross-references: EMBL:AL137556
 A:Experimental source: adult testis; clone DKFZp434i1120
 C:Genetics:
 A:Note: DKFZp434i1120.1

Query Match 47.8%; Score 100; DB 2; Length 380;
 Best Local Similarity 100.0%; Pred. No. 0.2; 0; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKKKKKK 45
 Db 355 KKKKKKKKKKKKKKKKKKKKK 374

RESULT 6
 S09388
 histone H1 - sea urchin (Parechinus angulosus)
 C:Species: Parechinus angulosus (angulate urchin)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
 C:Accession: S09388
 R:Hill, C.S.; Martin, S.R.; Thomas, J.O.
 EMBO J. 8, 2591-2599, 1989
 A:Title: A stable alpha-helical element in the carboxy-terminal domain of free and c
 A:Reference number: S09388; MUID:90060019
 A:Accession: S09388
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-206 <HIL>
 C:Superfamily: histone H1
 C:Keywords: chromosomal protein

Query Match 46.9%; Score 98; DB 2; Length 206;
 Best Local Similarity 55.8%; Pred. No. 0.19; 15; Indels 0; Gaps 0;
 Matches 24; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 EAAAAA...
 Db 140 KAAAKK... 182

RESULT 7
 S25194
 Zucotin - yeast (Saccharomyces cerevisiae)
 N:Alternate names: probable 2-DNA-binding protein; protein G9554; protein YGR285c
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
 C:Accession: S25194; S64620; S19066
 R:Zhang, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A.
 EMBO J. 11, 3787-3796, 1992
 A:Title: Zucotin, a putative 2-DNA binding protein in Saccharomyces cerevisiae.
 A:Reference number: S25194; MUID:93010971

Mon Jul 1 13:55:04 2002

us-09-461-684-4.rpr

Page 5

```

Qy      3  AAAAAEAAAAAAAAEAAAAAKKKKKKKK 34
        ||| ||| ||| ||| ||| ||| ||| |||
Db      176 AAAEARAAAAAEAAAAAKARAAAAEAKAKADK 207

```

Search completed: July 1, 2002, 06:28:12
Job time: 714 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:35 ; Search time 25.37 Seconds

(without alignments)
68.679 Million cell updates/sec

Title: US-09-461-684-4

Perfect score: 209
Sequence: 1 CEAFAAAEFAAAEFAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	49.3	248	1	H1_PANAN
2	96	45.9	433	1	ZOU1_YEAST
3	90	43.1	153	1	H2B4_CHIRE
4	89	42.6	207	1	H11_CABEL
5	86	41.1	347	1	TOLA_PSEAE
6	85	40.7	102	1	RL12_METVA
7	85	40.7	153	1	H2B3_CHIRE
8	85	40.7	240	1	H12_VOLCA
9	84	40.2	441	1	PHPA_PLACH
10	84	40.2	726	1	BRD3_HUACH
11	83	39.7	107	1	RL12_PIRAB
12	83	39.7	320	1	YD3_YEAST
13	83	39.7	421	1	TOLA_ECOLI
14	82	39.2	153	1	H2B1_CHIRE
15	82	39.2	474	1	CBF5_SCHPO
16	81	38.8	108	1	RL12_PPRHO
17	81	38.8	487	1	CBF5_PSPFU
18	80.5	38.5	534	1	NOP5_RAT
19	80.5	38.5	1002	1	IF2P_YEAST
20	80	38.3	111	1	RL12_AERPE
21	80	38.3	237	1	H1E_CHITE
22	80	38.3	508	1	NOG0_DROME
23	79.5	38.0	156	1	H2B2_CHIRE
24	79.5	38.0	511	1	NOP5_YEAST
25	79.5	38.0	2231	1	SEN1_YEAST
26	79	37.8	109	1	RLA1_YEAST
27	79	37.8	171	1	RLA1_TRYCR
28	79	37.8	260	1	H11_VOLCA
29	78.5	37.6	1391	1	MS12_DROHY
30	78	37.3	134	1	H1_EUPET
31	78	37.3	210	1	H1_LYPTI
32	78	37.3	218	1	H1R_CHICK
33	78	37.3	218	1	H12_RAT

34	78	37.3	218	1	H14_HUMAN
35	78	37.3	218	1	H14_MOUSE
36	78	37.3	372	1	TOLA_HAEIN
37	78	37.3	414	1	Y694_METVA
38	78	37.3	1220	1	IF2P_HUMAN
39	77.5	37.1	671	1	CHS5_YEAST
40	77	36.8	103	1	RL12_METTL
41	77	36.8	113	1	RLA1_DICDI
42	77	36.8	211	1	H13_MOUSE
43	77	36.8	220	1	H13_MOUSE
44	77	36.8	514	1	HRB_STRGR
45	77	36.8	862	1	VG01_HSVII

ALIGNMENTS

RESULT	1				
ID	H1_PANAN	STANDARD:	PRT:	248 AA.	
AC	P02256;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	Histone H1, gonadal.				
OS	Parechinus angulosus (Angulate sea urchin).				
OC	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;				
OC	Echinoidea; Euechinoidea; Echinacea; Echinoidae; Echinidae;				
OC	Parechinus.				
OX	NCBI_TaxID=7658;				
RN	[1]				
RP	SEQUENCE OF 1-84.				
RX	MEDLINE=80156831; PubMed=6767609;				
RA	Strickland W.N., Strickland M., de Groot P.C., von Holt C.,				
RA	Wittmann-Liebold B.;				
RT	"The primary structure of histone H1 from sperm of the sea urchin				
RT	Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the				
RT	protein and the sequence of amino acids in the four N-terminal				
RT	cytogen bromide peptides.";				
RL	Eur. J. Biochem. 104:559-566(1980).				
RN	[2]				
RP	SEQUENCE OF 80-248.				
RX	MEDLINE=80156832; PubMed=7363905;				
RA	Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,				
RA	Wittmann-Liebold B.;				
RT	"The primary structure of histone H1 from sperm of the sea urchin				
RT	Parechinus angulosus. 2. Sequence of the C-terminal CNBR peptide and				
RT	the entire primary structure.";				
RL	Eur. J. Biochem. 104:567-578(1980).				
CC	-1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF				
CC	NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.				
CC	-1- SUBCELLULAR LOCATION: Nuclear.				
CC	-1- TISSUE SPECIFICITY: SPERM.				
CC	-1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.				
DR	PIR; A02586; HSURIP.				
DR	HSSP; P02259; 1HST.				
DR	InterPro: IPR001386; Linker-histone.				
DR	InterPro: IPR003216; Linker-histone.				
DR	Pfam: PF00538; Linker-histone; 1.				
DR	ProDom: PD000373; Linker-histone; 1.				
DR	SMART; SM00526; H1S; 1.				
KW	Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;				
KW	Sperm.				
FT	SEQUENCE	144	144	K -> R.	
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144				
FT	144				
FT	K -> R.				
FT	SEQUENCE				
FT	248 AA; 26387 MW; 1B55B3F136541947 CRC64;				
FT	144</				

KW Bromodomain; Repeat: Nuclear protein.
 FT DOMAIN 56 115 BROMODOMAIN 1.
 FT DOMAIN 326 338 BROMODOMAIN 2.
 FT DOMAIN 487 535 LYS-RICH.
 FT DOMAIN 676 725 SER-RICH.
 FT CONFLICT 465 466 EL -> DV (IN REF. 2).
 SQ SEQUENCE 726 AA; 79541 MW; 64F526FC3C1033AA CRC64;

Query Match 40.2%; Score 84; DB 1; Length 726;
 Best Local Similarity 54.3%; Pred. No. 2;
 Matches 19; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 11 AA
 DB 472 AVHEQLAALSOAPVKNPKKKKKKKKKKKKKKKKKK 45

RESULT 11

RL12_PYRAB STANDARD; PRT; 107 AA.
 AC Q9UXS6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 50S ribosomal protein L12P.
 GN RPL12P OR PAB1168.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
 OX NCBI_TaxID=29292;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-ORSAY;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SEEKS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR ACCURATE TRANSLATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL_AJ248288; CAB50687.1; -
 DR InterPro; IPR001813; 60S_rribosomal.
 DR Pfam; PF00428; 60S_rribosomal; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 107 AA; 11300 MW; CDA23930BAD4ED CRC64;

Query Match 39.7%; Score 83; DB 1; Length 107;
 Best Local Similarity 44.2%; Pred. No. 0.34;
 Matches 19; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

QY 2 AA
 DB 55 EKAAAPVAAAPAAAPAEAPAEKKKEKKKEKKKEEVESEEE 97

RESULT 12

YD33_YEAST STANDARD; PRT; 320 AA.
 AC Q12117;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.
 GN YDR033W OR YD9673.03.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Arnold W., Becker A., Jaeger W., Kuester H., Nusebaumer B.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ARCHAEL OPSIN FAMILY. HSP30 SUBFAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z74329; CA98855.1; -
 DR EMBL; Z68196; CA92370.1; -
 DR SGD; S0002440; YDR033W.
 DR InterPro; IPR001425; Bac_rhodopsin.
 DR Pfam; PF01036; Bac_rhodopsin; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 63 83 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 168 188 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT DOMAIN 300 318 LYS-RICH.
 SQ SEQUENCE 320 AA; 36190 MW; 4311F64A6AA209F CRC64;

Query Match 39.7%; Score 83; DB 1; Length 320;
 Best Local Similarity 48.8%; Pred. No. 1.3;
 Matches 21; Conservative 4; Mismatches 14; Indels 4; Gaps 1;

QY 7 AA
 DB 277 AEAVAPAPAPVSPAPATPMLSKKKKKKKKKKKKKKSE 319

RESULT 13

TOLA_ECOLI STANDARD; PRT; 421 AA.
 AC P19934;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE TOLA protein.
 GN TOLA OR CIM OR EXCC OR LKY OR B0739.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-JM105;
 RX MEDLINE=90078104; Pubmed=2687247;
 RA Levgood S.K., Webster R.E.;
 RT "Nucleotide sequences of the tola and tolB genes and localization of their products, components of a multistep translocation system in

RT Escherichia coli."?
J. Bacteriol. 171:6600-6609(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97061202; Pubmed-9278503;
RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1245-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RX MEDLINE-97061202; Pubmed-8905232;
RA Osimine T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada Y., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-153(1996).
RN [4]
RN DOMAINS.
RP MEDLINE-91296736; Pubmed-2068069;
RA Levensgood S.K., Beyer W.F.Jr., Webster R.E.;
RT "TolA: a membrane protein involved in colicin uptake contains an
extended helical region.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN [5]
RN INTERACTION WITH PORINS.
RP MEDLINE-97133271; Pubmed-8978668;
RA Denouche R., Gavioit M., Benedetti H., Philippov A., Lazdunski C.,
RA Lioubes R.;
RT "TolA central domain interacts with Escherichia coli porins.";
RL EMBO J. 15:6408-6415(1996).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX MEDLINE-99332679; Pubmed-10404600;
RA Lubkowski J., Hennecke F., Plueckthun A., Mlodawer A.;
RT "Flamentous phage infection: crystal structure of TolA in complex
with its coreceptor, the C-terminal domain of TolA.";
RL Structure 7:11-722(1999).
CC -I- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION
OF BACTERIOPHAGE DNA.
CC -I- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPF, POE
AND LAMB.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28232; AAA24683.1; -
DR EMBL; AE000177; AAC73833.1; -
DR EMBL; D90713; BA835405.1; -
DR PIR; JY0057; JY0057.
DR PDB; 1TOL; 2O-MAY-99.
DR Ecogene; EGI1007; tolA.
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW Repeat; Inner membrane; 3d-structure; Complete proteome.
FT DOMAIN 1 13
FT DOMAIN 14 34
FT POTENTIAL.

```

FT DOMAIN 35 421 PERPLASMIC (POTENTIAL).
FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).
FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
FT DOMAIN 224 278 10 X TANDED REPEATS OF [ED]-K(1,2)-
FT A(2,4)'.
SQ SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;

Query Match 39.7%; Score 83; DB 1; Length 421;
Best Local Similarity 47.18; Pred. No. 1.6; Indels 8; Gaps 1;
Matches 24; Conservative 5; Mismatches 14;

QY 3 AAAAAAEEEEAAAA-----AAEAAAAKKKKKKKKKKKKKKKKKK 45
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 AKAAAEAGCKRAEAAAALKKRAEAAEAAAEAAEAKRAAEAEAKKAEEAEKK 226

RESULT 14
ID H2B1_CHLRE STANDARD; PRT; 153 AA.
AC P50565;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Histone H2B-I.
GN H2B.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RX MEDLINE=96017782; PubMed=7479007;
RA Walther Z., Hall J.L.;
RT "The uni-chromosome of Chlamydomonas: histone genes and nucleosome
   structure."
RL Nucleic Acids Res. 23:3756-3763(1995).
CC -1 SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
   H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1 SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L41841, AAA99967.1; -
DR InterPro: IPR000558; Histone_H2B.
DR InterPro: IPR000166; Histone_core.
DR Pfam: PF00125; histone; 1.
DR PRINTS: PR00621; HISTONEH2B.
DR PRODOM: PD000497; Histone_H2B; 1.
DR SMART: SM00427; H2B; 1.
DR PROSITE: PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW Multigene family.
SQ SEQUENCE 153 AA; 16727 MW; BAD7F264B4544EF CRC64;

Query Match 39.2%; Score 82; DB 1; Length 153;
Best Local Similarity 48.9%; Pred. No. 0.85;
Matches 23; Conservative 2; Mismatches 16; Indels 6; Gaps 1;

QY 5 AAAAAAEEEEAAAAAAXKKKKKKKKKK-----KKKKKKKK 45
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14 AGAEAPAKAERKPKAEKAAKKRAKKPKSKAAKEPVGSGEKKKKKK 60

```

```
RESULT 15
CBF5_SCHPO STANDARD: PRT: 474 AA.
ID CBF5_SCHPO
AC 014007;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Centromere/microtubule binding protein cbf5 (centromere-binding factor
DE 5) (Nucleolar protein cbf5).
GN CBF5 OR SPAC29A4.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z97210; CAB10131.1; -
DR InterPro: IPR002478; PUA.
DR InterPro: IPR002501; TRUB_N.
DR Pfam: PF01472; PUA; 1.
DR SMART: SM00359; PUA; 1.
DR Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
KW DOMAIN 434 468 7 X 3 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 443 445 1. K-K-E.
FT REPEAT 450 452 2.
FT REPEAT 454 456 3.
FT REPEAT 457 459 4.
FT REPEAT 460 462 5.
FT REPEAT 463 465 6.
FT REPEAT 466 468 7.
SQ SEQUENCE 474 AA; 53110 MM; B8C9896C5FAEB71 CRC64;
```

```
Query Match 39.28; Score 82; DB 1; Length 474;
Best Local Similarity 34.18; Pred. No. 2.1;
Matches 20; Conservative 5; Mismatches 8; Indels 4; Gaps 1;
```

```
OY 8 EAAAAAEEAAAEAAAKKKKKKKKKKKKKKKKK 44
||| | | | | | | | | | | | | | | | |
DB 440 EAAKKEKKRKKK---AKKEKKKKKKKKKKKK 472
```

Search completed: July 1, 2002, 06:31:35
Job time: 672 sec

Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Metkulov G., Mishina N.V., Kobayashi C., Morris J., Moshel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shee B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stadler A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter A., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AF036698; AAF54888.2; -
DR FlyBase; FBgn0038108; CG7518.
DR InterPro; IPR001005; Myb_DNA_bind.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 2451 AA; 266959 MW; 088A2293F27481E2 CRC64;

Query Match 56.5%; Score 118; DB 5; Length 2451;
Best Local Similarity 75.7%; Pred. No. 0.014;
Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 3 AAAAAAAAAAAAAAAAAAKKKKKKKKKKKKK 39
DB 1374 AAAAAAAAAAAAAAAAAEAAEQAKKKKKQAKK 1410

RESULT 2
ID 035807 PRELIMINARY; PRT; 129 AA.
AC 035807;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.
GN MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EPIDIDYMIS;
RX MEDLINE=98172708; PubMed=9511718;
RA Proels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
species, during in vitro angiogenesis";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL; Y08769; CAA70022.1; -
DR InterPro; IPR000719; Euk_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KM ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BEEZEDB4 CRC64;

Query Match 50.2%; Score 105; DB 11; Length 129;
Best Local Similarity 95.5%; Pred. No. 0.013; 0; Indels 0; Gaps 0;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 24 AAKKKKKKKKKKKKKKKKKKKKK 45
DB 83 AAKKKKKKKKKKKKKKKKKKKKK 104

RESULT 3
O9H607

ID 09H607 PRELIMINARY; PRT; 720 AA.
AC 09H607;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CDNA: FLJ21979 F1S, CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kodatate N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Oshayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK025632; BAB15196.1; -
DR NON_TER 720
FT 720
SQ SEQUENCE 720 AA; 84029 MW; A86586FEAA953D0B CRC64;

Query Match 50.2%; Score 105; DB 4; Length 720;
Best Local Similarity 61.8%; Pred. No. 0.061; 7; Indels 0; Gaps 0;
Matches 21; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 12 AAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKK 45
DB 678 AKKINSNDIVSISKKKKKKKKKKKKKKKKKK 711

RESULT 4
ID 09L82 PRELIMINARY; PRT; 467 AA.
AC 09L82;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CBP5.
OS Euglena gracilis.
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
OX NCBI_TaxID=3039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STREPTOMYCIN-BLEACHED STRAIN;
RX MEDLINE=20330353; PubMed=10871366;
RA Watanabe Y., Gray M.W.;
RT "Evolutionary appearance of genes encoding proteins associated with
box H/ACA snRNAs: CBP5 in Euglena gracilis, an early diverging
eukaryote, and candidate Garlp and Nop1op homologs in
archaeobacteria";
RL Nucleic Acids Res. 28:2342-2352(2000).
DR EMBL; AF234319; AAF77119.1; -
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; TRUB_N.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TRUB_N; 1.
DR SMART; SM00359; PUA; 1.
SQ SEQUENCE 467 AA; 52385 MW; 401089B66507BA7B CRC64;

Query Match 48.8%; Score 102; DB 10; Length 467;
Best Local Similarity 55.8%; Pred. No. 0.075; 14; Indels 0; Gaps 0;
Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 EAAAAAAAAAAAAAAAAAKKKKKKKKKKKKKKK 44
DB 420 EELAYAAAKRREREAAGDEKAKKAKKKKKKKKK 462

RESULT 5
O9LXR2 PRELIMINARY; PRT; 517 AA.
ID O9LXR2

AC Q9LXR2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHETICAL 59.7 KDA PROTEIN.
GN T20N10-250.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.,
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL353032; CAB8307.1; -
DR InterPro: IPR001810; F-box.
DR Pfam: PF00646; F-box; 1.
DR SMART: SM00256; FBOX; 1.
DR PROSITE: PS50181; FBOX; 1.
DR Hypothetical protein.
SQ SEQUENCE 517 AA; 59689 MW; EC6D957D01F86E70 CRC64;
Query Match 48.8%; Score 102; DB 10; Length 517;
Best Local Similarity 80.8%; Pred. No. 0.082; Mismatches 5; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 20 EAAAAA
DB 440 ERVGGKKKKKKKKKKKKKKKKKKKK 465
RESULT 6
ID 014347 PRELIMINARY; PRT; 354 AA.
AC 014347;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE ZUOTIN-LIKE PROTEIN C30D10.01 (FRAGMENT).
GN SP3C30D10.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STAINE-972;
RA Wood V., Rajandream M.A., Barrell B.G., Lauber J., Hilbert H.,
RA Duesterhoeft A.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN CHROMOSOME
CC ORGANIZATION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
DR EMBL: Z97992; CAB10796.1; -
DR HSSP: P25685; 1HD;
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ; 1.
DR SMART: SM00271; DnaJ; 1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Hypothetical protein; Chaperone; DNA-binding; Nuclear protein.
FT NON-TER 1
FT DOMAIN 10 81 DnaJ-LIKE.
FT 217 267 ALA/LYS-RICH.

SQ SEQUENCE 354 AA; 40290 MW; 6071B58A3B60F558 CRC64;
Query Match 48.6%; Score 101.5; DB 3; Length 354;
Best Local Similarity 56.0%; Pred. No. 0.064; Mismatches 12; Indels 7; Gaps 1;
Matches 28; Conservative 3; Mismatches 12; Indels 7; Gaps 1;
QY 2 EAAAAA-----EAAAAA
DB 221 EAAAAAOKKKEEERAAAEAAAKASAAANKAKEDKKAKGRKKYK 270
RESULT 7
ID 0951V6 PRELIMINARY; PRT; 531 AA.
AC 0951V6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Tero K., Sugano S.;
RT Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries.
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB071085; BAB64479.1; -
DR Eukaryotic protein.
KW NON-TER 531
FT NON-TER 531
SQ SEQUENCE 531 AA; 61389 MW; B5596BAF5CDD60C CRC64;
Query Match 48.3%; Score 101; DB 6; Length 531;
Best Local Similarity 95.2%; Pred. No. 0.1; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 25 AKKKKKKKKKKKKKKKKKKKKK 45
DB 501 SKKKKKKKKKKKKKKKKKKKKK 521
RESULT 8
ID 09P529 PRELIMINARY; PRT; 128 AA.
AC 09P529;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 15.2 KDA PROTEIN.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hohnsels J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL356815; CAB92638.2; -
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Query Match 47.8%; Score 100; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45
DB 71 KKKKKKKKKKKKKKKKK 90

RESULT 9
Q9H5V6 PRELIMINARY; PRT; 168 AA.
AC Q9H5V6;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuyoshi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK026629; BAB15513.1; -
FT NON_TER 168
SQ SEQUENCE 168 AA; 19549 MW; A19DBD195F8A1A90 CRC64;

Query Match 47.8%; Score 100; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45
DB 140 KKKKKKKKKKKKKKKKK 159

RESULT 10
Q64075 PRELIMINARY; PRT; 215 AA.
AC Q64075;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE NUCLEOPORIN P62 HOMOLOG PROTEIN (FRAGMENT).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Solurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:95151924; Pubmed-7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
in the germ cells of rat testis."
RL Biol. Reprod. 51:1022-1030(1994).
DR EMBL: S75997; AAB3384.1; -
FT NON_TER 1
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

Query Match 47.8%; Score 100; DB 11; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45
DB 35 KKKKKKKKKKKKKKKKK 54

RESULT 11
Q9LG29 PRELIMINARY; PRT; 260 AA.
AC Q9LG29;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE: FID9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLLEMBIA;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002460; BAA97098.1; -
DR InterPro: IPR001386; Linker_histone.
DR PRINTS: PRO0624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E294C83131143 CRC64;

Query Match 47.8%; Score 100; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 0.065;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45
DB 7 KKKKKKKKKKKKKKKKK 26

RESULT 12
Q9NT34 PRELIMINARY; PRT; 380 AA.
AC Q9NT34;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE HYPOTHEICAL 42.7 KDA PROTEIN (FRAGMENT).
GN DKFZP434I1120.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL137556; CAB70810.1; -
RT Hypothetical protein.
FT NON_TER 380
SQ SEQUENCE 380 AA; 42689 MW; 67F5DD010346AFB CRC64;

Query Match 47.8%; Score 100; DB 4; Length 380;
Best Local Similarity 100.0%; Pred. No. 0.092;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KKKKKKKKKKKKKKKKK 45
DB 355 KKKKKKKKKKKKKKKKK 374

RESULT 13

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:24 ; Search time 98.97 Seconds

(without alignments)
50.503 Million cell updates/sec

Title: US-09-461-684-4

Perfect score: 209
Sequence: 1 CEAATAAAATAAAATAAAE.....KKKKKKKKKKKKKKKKKK 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A_Geneseq_032802:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	209	100.0	45	21 AAB13783
2	119	56.9	36	21 AAB23585
3	119	56.9	630	21 AAB23591
4	119	56.9	640	21 AAB23593
5	118	56.5	2451	22 ABB17574
6	115.5	55.3	123	22 AAO01368
7	115	55.0	113	22 AAO03703
8	111	53.1	63	22 AAT18275
9	110	52.6	141	22 AAG26718
10	109	52.2	25	21 AAB13781
11	109	52.2	59	21 AAY98495

Result No.	Score	Query Match Length	DB ID	Description
12	109	52.2	59	21 AAY99040
13	109	52.2	59	22 AAU04285
14	109	52.2	59	22 AAB45848
15	109	52.2	74	22 AAO03278
16	109	52.2	75	22 AAO08943
17	109	52.2	112	22 AAO02477
18	109	52.2	150	22 AAU18240
19	108	51.7	80	22 AAO12105
20	108	51.7	126	22 AAO02945
21	107	51.2	48	22 AAO08707
22	107	51.2	63	22 AAU18192
23	107	51.2	83	22 AAO02961
24	106	50.7	26	22 AAO08995
25	106	50.7	57	22 AAG26720
26	106	50.7	64	22 AAO03024
27	106	50.7	74	22 AAU18205
28	106	50.7	74	22 AAU18277
29	106	50.7	128	21 AAY86248
30	106	50.7	272	22 AAU23799
31	105	50.2	31	22 AAB59105
32	105	50.2	31	22 AAO08450
33	105	50.2	62	20 AAY12950
34	105	50.2	62	22 AAO11165
35	105	50.2	66	22 AAU21948
36	105	50.2	66	22 AAO07505
37	105	50.2	72	22 AAU18178
38	105	50.2	83	22 AAG73729
39	105	50.2	84	22 AAU18262
40	105	50.2	122	22 AAO10995
41	105	50.2	155	22 AAG75215
42	105	50.2	175	11 AAR08261
43	105	50.2	630	21 AAB23592
44	105	50.2	640	21 AAB23594
45	104	49.8	28	22 AAO08817

ALIGNMENTS

Result No.	Score	Query Match Length	DB ID	Description
1	209	100.0	45	21 AAB13783
2	119	56.9	36	21 AAB23585
3	119	56.9	630	21 AAB23591
4	119	56.9	640	21 AAB23593
5	118	56.5	2451	22 ABB17574
6	115.5	55.3	123	22 AAO01368
7	115	55.0	113	22 AAO03703
8	111	53.1	63	22 AAT18275
9	110	52.6	141	22 AAG26718
10	109	52.2	25	21 AAB13781
11	109	52.2	59	21 AAY98495

ID AAB23593 standard; Protein; 640 AA.
 XX
 AC AAB23593;
 XX
 DT 11-JAN-2001 (first entry)
 XX
 DE Modified fibre protein encoded in pWE6.7R-F/ask21MSHB.
 XX
 KM Viral vector; melanocyte-stimulating hormone receptor; MSH; cytostatic;
 XX tumour; malignant melanoma; fibre protein.
 OS Synthetic.
 XX
 PN WO200050618-A1.
 XX
 PD 31-AUG-2000.
 XX
 PF 24-FEB-2000; 2000WO-JP01069.
 XX
 PR 24-FEB-1999; 99JP-0093263.
 XX
 PA (NICA-) JAPANESE FOUND CANCER RES.
 XX
 PI Hamada H;
 XX
 DR WPI; 2000-549414/50.
 XX
 DR N-PSDB; AAA93825.
 XX
 PT Virus vector useful in diagnosis and treatment of tumors particularly
 PT malignant melanoma; constructed by fusing viral protein with ligand
 PT binding specifically to melanocyte-stimulating hormone receptor
 XX
 PS Claim 24; Page 131-136; 145pp; Japanese.
 XX
 CC This invention relates to a viral vector constructed by the fusion of a
 CC viral protein with a ligand which binding specifically to the
 CC melanocyte-stimulating hormone (MSH) receptor. The vector contains one
 CC of four linkers represented by sequences AAA93815-A93818 and
 CC AAB23583-B23586, and DNA encoding a fibre protein selected from those
 CC represented by AAA93819-A93826 and AAB23587-B23594. The vector has
 CC cytostatic activity, and can be used for gene therapy and in the
 CC diagnosis and treatment of tumours, particularly malignant melanomas. The
 CC present sequence represents a modified fibre protein used in the
 CC construction of the vector.
 CC
 SQ Sequence 640 AA;
 XX
 Query Match 56.9%; Score 119; DB 21; Length 640;
 Best Local Similarity 70.6%; Pred. No. 0.00016;
 Matches 24; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 12 AAEAAAAAEAAA 45
 Db 583 saasasaaagsakkkkkkkkkkkkkkkkkkkkkkkkkkkkk 616
 RESULT 5
 ABB71574
 ID ABB71574 standard; Protein; 2451 AA.
 XX
 AC ABB71574;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 41514.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.

XX
 PD 27-SEP-2001.
 XX
 AC 23-MAR-2001; 2001WO-US09231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL15677.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 41514; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 2451 AA;
 XX
 Query Match 56.5%; Score 118; DB 22; Length 2451;
 Best Local Similarity 75.7%; Pred. No. 0.00072;
 Matches 28; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 AAAAAAEAAAAAEAAA 39
 Db 1374 aaaaaaataaaaaaataaaaaaataaaagkxkxkxkxgaxk 1410
 RESULT 6
 AA001368
 ID AA001368 standard; Protein; 123 AA.
 XX
 AC AA001368;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 15260.
 XX
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 XX
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.

OY 1 CEAHAADPAHAHAHAHAHAHAHAHA 25
 ID 1 CEAHAADPAHAHAHAHAHAHAHAHA 25

RESULT 11

AA98495
 ID AAY98495 standard; Peptide: 59 AA.

AA98495;

31-JUL-2000 (first entry)

Nuclear ligand used in nucleic acid transporter system.

Transporter system; nucleic acid delivery; gene therapy; cancer;

cardiogenesis; cardiovascular disease; infection.

Synthetic.

US6033884-A.

07-MAR-2000.

14-DEC-1993; 93US-0167641.

20-MAR-1992; 92US-0855389.

19-MAR-1993; 93WO-US02725.

(BAYU) BAYLOR COLLEGE MEDICINE.

Gottchalk S, Sparrow J, Cristiano RJ, Woo SLC, Smith LC;

WPI; 2000-281993/24.

System for transporting nucleic acid into cells; useful e.g. in gene therapy and for generating transgenic animals; comprises binding agent linked to nucleic acid, surface ligand and lytic agent.

Claim 16; Column 123-124; 108pp; English.

The present invention relates to a transporter system for delivering nucleic acid to a cell. The system comprises a nucleic acid binding complex, consisting of a binding molecule bonded non-covalently to the nucleic acid, and covalently to a surface ligand, and a lytic agent. The binding molecule is spermine or a spermidine derivative. Nucleotide sequences AA36633-A36652 and peptide sequences AAY98495-Y98500 are used in the construction of the transporter system of the invention. The transporter system is used in gene therapy, particularly to deliver nucleic acids to hepatocytes, muscle cells or bone forming cells, e.g. for treating cardiovascular disease, cancer, and infection. The transporter systems are also used to create transgenic animals (as models for human carcinogenesis or disease or for drug testing). Other uses include transforming cells to produce proteins, or transfecting cells in vitro to study the function of the nucleic acid. The use of a surface ligand allows specific targeting of selected cells and tissues. The lytic agent provides for release of the nucleic acid into the cellular interior, from endosomes, without requiring endosomal or lysosomal degradation.

Sequence 59 AA;

Query Match 52.2%; Score 109; DB 21; Length 59;
 Best Local Similarity 88.5%; Pred. No. 0.00016;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 20 EAAAAAHHHHHHHHHHHHHHHHHHH 45
 ID 12 eapyaakkkkkkkkkkkkkkkkkkkk 37

RESULT 12

AA59040
 ID AAY59040 standard; peptide: 59 AA.

AA59040;

07-MAR-2000 (first entry)

Nuclear ligand for transporting nucleic acid to the nucleus.

Nucleic acid transport system; NTS; cell surface receptor; cytosol;

nuclear membrane; lysis moiety; transgenic animal; human disease;

nucleic acid delivery; cancer.

Synthetic.

US5994109-A.

30-NOV-1999.

03-JUN-1995; 95US-0460890.

14-DEC-1993; 93US-0167641.

20-MAR-1992; 92US-0855389.

19-MAR-1993; 93WO-US02725.

14-DEC-1993; 93US-0167641.

(BAYU) BAYLOR COLLEGE MEDICINE.

Woo SLC, Cristiano RJ, Gottchalk S, Sparrow J, Smith LC;

WPI; 2000-038262/03.

Nucleic acid transport system, useful for creating transgenic animals for assessing human disease such as cancer in an animal model.

Disclosure; Columns 119-122; 107pp; English.

The invention relates to a nucleic acid transport system (NTS) for delivering nucleic acid into a cell. The NTS contains but is not limited to 5 components: (a) the nucleic acid or a macromolecule to be delivered; (b) a moiety that recognizes and binds to a cell surface receptor or antigen or is capable of entering a cell through cytosol; (c) a nucleic acid or macromolecular molecule binding moiety; (d) a moiety that is capable of moving or initiating movement through a nuclear membrane; and/ or (e) a lysis moiety that enables the transport of the entire complex from the cell surface directly into the cytoplasm of the cell. The NTS delivers nucleic acid into the cellular interior as well as the nucleus of specific cells. The NTS can be used to treat disorders by targeting specific nucleic acid accordingly. The NTS can also be used to create transgenic animals for assessing human disease, such as cancer, in an animal model. The NTS can be used in vitro with tissue culture cells which allows the role of various nucleic acids to be studied by targeting specific expression into specifically targeted tissue culture cells. The lysis agent within the NTS avoids the problem of endosomal/lysosomal degradation.

Sequence 59 AA;

Query Match 52.2%; Score 109; DB 21; Length 59;
 Best Local Similarity 88.5%; Pred. No. 0.00016;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 20 EAAAAAHHHHHHHHHHHHHHHHHHH 45
 ID 12 eapyaakkkkkkkkkkkkkkkkkkkk 37

RESULT 13
 AA004285
 ID AA004285 standard; Peptide: 59 AA.
 XX XX
 AC AA004285;

XX 23-OCT-2001 (first entry)
DT
XX
XX Nuclear ligand #2 used in nucleic acid transporter system.
DE
XX
XX Nucleic acid transporter; cytosols; ligand; lysis agent; spacer molecule;
KW gene therapy; hepatocyte; muscle; bone forming cell.
KV
XX
XX Synthetic.
OS
XX US6177554-B1.
PN
XX 23-JAN-2001.
PD
XX
XX 05-JUN-1995; 95US-0462040.
PF
XX
XX 14-DEC-1993; 93US-0167641.
PR 20-MAR-1992; 92US-0855389.
PR 19-MAR-1993; 93WO-US02725.
XX
XX (BAYU) BAYLOR COLLEGE MEDICINE.
PA
XX
XX Woo SLIC, Smith LC, Cristiano RJ, Gottchalk S, Sparrow J;
PI
XX WPI: 2001-365933/38.
DR
XX
XX Nucleic acid transport system, useful for creating transgenic animals
PT for assessing human disease such as cancer in an animal model .
PS
XX
XX Disclosure: Column 8; 111pp; English.
CC
XX The sequence represents the nuclear ligand, #2, used in a nucleic acid
CC transporter system. The nucleic acid transporter system uses nucleic acid
CC binding complexes containing surface ligands which are capable of binding
CC to a cell surface receptor and entering the cell through cytosols. The
CC compounds of the invention are either ligands, binding molecules (surface
CC ligands), lysis agents, spacer molecules or their intermediates. The
CC ligands, binding molecules, lysis agents and spacer molecules are used in
CC nucleic acid transporter systems to deliver nucleic acid into specific
CC cells e.g. in gene therapy to deliver nucleic acid into hepatocytes,
CC muscle cells or bone forming cells.
CC
XX
XX Sequence 59 AA:
SQ

Query Match 52.2%; Score 109; DB 22; Length 59;
Best Local Similarity 88.5%; Pred. No. 0.00016;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 EAAAKKKKKKKKKKKKKKKKKKKKK 45
DB 12 eapYakKKKKKKKKKKKKKKKKKK 37

RESULT 14
AAB45848
ID AAB45848 standard; Protein: 59 AA.
XX
XX AAB45848;
AC
XX
XX 21-MAR-2001 (first entry)
DT
XX
XX Nucleic acid transporter system peptide ligand SEQ ID NO 60.
DE
XX
XX Nucleic acid delivery; nucleic acid transporter system; hormone; enzyme;
KW growth factor; clotting factor; apolipoprotein; receptor; drug; oncogene;
KW tumor antigen; tumor suppressor; viral antigen; parasitic antigen;
KW bacterial antigen.
XX
XX (unidentified.
OS
XX US6150168-A.
PN
XX

PD 21-NOV-2000.
XX
XX 05-JUN-1995; 95US-0460971.
PF
XX
XX 14-DEC-1993; 93US-0167641.
PR 20-MAR-1992; 92US-0855389.
PR 19-MAR-1993; 93WO-US02725.
XX
XX (BAYU) BAYLOR COLLEGE MEDICINE.
PA
XX
XX Gottchalk S, Sparrow J, Cristiano RJ, Smith LC, Woo SLIC;
PI
XX WPI: 2001-049093/06.
DR
XX
XX Nucleic acid transporter system for delivering nucleic acid into a
PT cell, useful for delivering proteins and polypeptides to cells,
PT including growth factors, enzymes, hormones, and tumor suppressors
XX
XX Disclosure: Column 123-124; 105pp; English.
CC
XX This invention describes a novel system (I) for delivering a nucleic acid
CC to a cell, comprising a binding complex comprising a ligand binding
CC molecule noncovalently bound to a nucleic acid and covalently linked to a
CC surface ligand, and a second binding complex comprising a second binding
CC molecule noncovalently bound to a nucleic acid and covalently linked to a
CC nuclear ligand. The complexes are simultaneously bound to the nucleic
CC acid. The nucleic acid transporter system can also be used in a method
CC for the in vivo targeting of the insertion of DNA into a cell. It can
CC also be used in processes for producing transformed cell lines. The
CC system can be used to deliver a variety of proteins and polypeptides,
CC such as hormones, growth factors, enzymes, clotting factors,
CC apolipoproteins, receptors, drugs, oncogenes, tumor antigens, tumor
CC suppressors, viral antigens, parasitic antigens, and bacterial antigens.
CC The transporter system uses lysis agents to overcome the problems of
CC endosomal/lysosomal degradation seen with prior art systems.
CC
XX
XX Sequence 59 AA:
SQ

Query Match 52.2%; Score 109; DB 22; Length 59;
Best Local Similarity 88.5%; Pred. No. 0.00016;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 20 EAAAKKKKKKKKKKKKKKKKKKKKK 45
DB 12 eapYakKKKKKKKKKKKKKKKKKK 37

RESULT 15
AAO03278
ID AAO03278 standard; Protein: 74 AA.
XX
XX AAO03278;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 17170.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX WO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US04927.
PF
XX
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
PR

XX
PA (HYSE-) HYSEQ INC.
XX

PI Tang YT, Liu C, Drmanac RT;
YY

DR WPI; 2001-514838/56.
DR N-PSDB; AA183209.

DR N-PSDB; AA183209.

PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX

PT disorders -

disorders -
PT
vxy

PS Claim 20; SEQ ID NO 17170; 1399pp + Sequence Listing; English
XX

The invention relates to human polypeptides (AA179941-AA19341) and the encoded proteins (AA000010-AA013910) that exhibit activity eliciting to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polypeptides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activating/inhibiting activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/patdb/published_pat_sequences.

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

CC at ftp.wipo.int/pub/published_pct_sequences.
YY

Sequence 74 AA;

Query Match
Best Local

52.28; Score 109; DB 22; Length 74;
77.49; Prod No 0 0003.

```

1000 Local similarity 77.48; Fied: NO: 0.00002;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      15 AAAAAEAAAAAAAAKKKKKKKKKKKKKKKKKKKK 45
          | | | | | | | | | | | | | | | |
Db      36 algalravaqkkkkkkkkkkkkkkkkkkkk 66

```

Db 36 aligalravaqkkkkkkkkkkkkkkkkkkkkk 66

Search completed: July 1, 2002, 06:19:25
Job time: 512 sec

Job time: 512 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:28:12 ; Search time 49.45 Seconds

(without alignments)
85,499 Million cell updates/sec

Title: US-09-461-684-5

Perfect score: 243

Sequence: 1 GGLFGAIGFIENGWEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_711**
2: PIR1**
3: PIR2**
4: PIR3**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	550	1 HMTVS2	hemagglutinin prec
2	134	55.1	550	1 HMTVS3	hemagglutinin prec
3	134	55.1	550	1 HMTV77	hemagglutinin prec
4	134	55.1	550	1 HMTV80	hemagglutinin prec
5	134	55.1	550	1 HMTV33	hemagglutinin prec
6	134	55.1	550	1 HMTV89	hemagglutinin prec
7	134	55.1	550	1 HMTV21	hemagglutinin prec
8	134	55.1	550	1 HMTV98	hemagglutinin prec
9	134	55.1	550	1 HMTV15	hemagglutinin prec
10	134	55.1	550	2 JQ1153	hemagglutinin prec
11	134	55.1	550	2 JQ1154	hemagglutinin prec
12	134	55.1	550	2 JQ1155	hemagglutinin prec
13	134	55.1	550	1 HMTV14	hemagglutinin prec
14	134	55.1	550	1 HMTV14	hemagglutinin prec
15	134	55.1	550	1 HMTV14	hemagglutinin prec
16	134	55.1	550	1 HMTV14	hemagglutinin prec
17	134	55.1	550	1 A45591	hemagglutinin prec
18	134	55.1	550	2 S22013	hemagglutinin prec
19	133	54.7	561	1 HMTV49	hemagglutinin prec
20	133	54.7	561	1 HMTV49	hemagglutinin prec
21	132	54.3	565	1 HMTVE1	hemagglutinin prec
22	132	54.3	565	1 HMTVE3	hemagglutinin prec
23	132	54.3	566	1 HMTVE6	hemagglutinin prec
24	132	54.3	567	1 HMTV17	hemagglutinin prec
25	131	53.9	560	2 S38637	hemagglutinin - in
26	131	53.9	560	1 HMTV86	hemagglutinin prec
27	131	53.9	560	1 HMTV77	hemagglutinin prec
28	131	53.9	565	1 HMTVE2	hemagglutinin prec
29	131	53.9	565	1 HMTVE4	hemagglutinin prec

30	131	53.9	565	1 HMTVE5	hemagglutinin prec
31	131	53.9	565	1 HMTVE6	hemagglutinin prec
32	131	53.9	565	1 HMTVE7	hemagglutinin prec
33	131	53.9	565	1 HMTVE8	hemagglutinin prec
34	131	53.9	565	1 HMTVE9	hemagglutinin prec
35	131	53.9	565	1 HMTVE9	hemagglutinin prec
36	131	53.9	565	1 HMTVE9	hemagglutinin prec
37	131	53.9	565	1 HMTVE9	hemagglutinin prec
38	131	53.9	570	2 S22014	hemagglutinin prec
39	131	53.9	570	2 S22015	hemagglutinin prec
40	131	53.9	570	2 S22016	hemagglutinin prec
41	131	53.9	570	2 S22017	hemagglutinin prec
42	131	53.9	570	2 S22018	hemagglutinin prec
43	131	53.9	570	2 S22019	hemagglutinin prec
44	131	53.9	570	2 S22020	hemagglutinin prec
45	131	53.9	570	2 S22021	hemagglutinin prec

ALIGNMENTS

RESULT 1
HMTVS2 hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C:Species: Influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: A29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China
A:Reference number: A94370; MUID:88101364
A:Accession: A29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19056; NID:g324208
A:Note: The sequence in Genbank entry FLAHAP, release 106, (PID:g324209) differs fro
C:Genetics:
A:Map position: segment 4
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HAI #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8-22-38-165-285-483/Binding site: carbohydrate (asn) (covalent) #status predicted
F:14-466-52-277-64-76-139-473-201-305/Disulfide bonds: #status predicted
F:535/546-349/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match
Best Local Similarity 55.1%; Score 134; DB 1; Length 550;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGLFGAIGFIENGWEGMIDGWTG 24
DB 330 GGLFGAIGFIENGWEGMIDGWTG 352

RESULT 2
HMTVS3 hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C:Species: Influenza A virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C:Accession: B29971
R:Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A:Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China
A:Reference number: A94370; MUID:88101364
A:Accession: B29971
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M19057; NID:g324210
A:Note: The sequence in Genbank entry FLAHAP, release 106, (PID:g324211) differs fro
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 3
HMTV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A27813
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: A27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16737; NID:g324081; PIDN:AAA43145.1; PID:g324082
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 4
HMTV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: B27813
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: B27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16738; NID:g324083
A:Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C:Genetics:

A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 5
HMTV73
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: C27813
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: C27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domain: transmembrane #status predicted <TM>
F:8-22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
DB 330 GLFGAIGFIENGWEGMIDGMYG 352

RESULT 6
HMTV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: D27813
R:Kida, H.; Kawoka, Y.; Naewe, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: D27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
C:Genetics:
A:Map position: segment 4

C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:520-536/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
|||||
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 7

HAIIV21
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: E27813
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: E27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16741; NID:9324089
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:520-536/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
|||||
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 8

HAIIV98
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C:Accession: F27813
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: F27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16742; NID:9324091
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:520-536/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
|||||
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 9

HAIIV15
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: G27813
R:Kida, H.; Kawoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A:Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A:Reference number: A94363; MUID:87265458
A:Accession: G27813
A:Molecule type: genomic RNA
A:Residues: 1-550 <KID>
A:Cross-references: GB:M16743; NID:9324093; PIDN:AAA3149.1; PID:9324094
C:Genetics:
A:Map position: segment 4
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:520-536/Product: hemagglutinin HA2 #status predicted <HA2>
F:520-536/Domains: transmembrane #status predicted <TM1>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F:539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1% Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
|||||
DB 330 GLFGAIGFIENGWEGMIDGWTG 352

RESULT 10

JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: JQ1153
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian
A:Reference number: JQ1153; MUID:91341491
A:Accession: JQ1153
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00929; NID:9221279; PIDN:BA00769.1; PID:9221280
A:Note: the authors translated the codon GGG for residue 218 as Glu
A:Note: residues 528-532 are not shown in this publication
C:Superfamily: influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>

F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
Db 330 GLFGAIGFIENGWEGMIDGWYG 352

RESULT 11

hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01154
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491
A:Accession: J01154
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00930; NID:q221273; PIDN:AAA00770.1; PID:q221274
A:Note: the authors translated the codon GGG for residue 218 as Glu
C:Note: residues 528-532 are not shown in this publication
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
Db 330 GLFGAIGFIENGWEGMIDGWYG 352

RESULT 12

hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C:Accession: J01155
R:Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A:Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A:Reference number: J01153; MUID:91341491
A:Accession: J01155
A:Molecule type: genomic RNA
A:Residues: 1-550 <YAS>
A:Cross-references: GB:D00931; NID:q221277; PIDN:AAA00771.1; PID:q221278
A:Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538
C:Note: residues 528-532 are not shown in this publication
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; homotrimer
F:1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F:330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F:8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 2; Length 550;
Best Local Similarity 100.0%; Pred. No. 3,1e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
Db 330 GLFGAIGFIENGWEGMIDGWYG 352

RESULT 13

hemagglutinin precursor - influenza A virus
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C>Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C:Accession: A93705; A93233; A04051; A93231; A94441
R:Both, G.W.; Sleigh, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A:Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza A virus
A:Reference number: A93705; MUID:81053698
A:Accession: A93705
A:Molecule type: genomic RNA
A:Residues: 1-566 <BOT>
A:Cross-references: GB:V01103
A:Experimental source: strain A/NT/60/68/29C
A:Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/Ridgely, T.A.; Maiz, C.W.
FBS Lett. 110, 181-183, 1980
A:Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A:Reference number: A91276; MUID:80179105
A:Contents: annotation; disulfide bonds
R:Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A:Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes
A:Reference number: A93233; MUID:81030852
A:Accession: A93233
A:Molecule type: genomic RNA
A:Residues: 1-24, 'S', '26', 'D', '28-159', 'G', '161-197', 'I', '199-241', 'L', '243-249' <GET>
A:Experimental source: strain X-31[H3]
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:536-552/Domain: transmembrane #status predicted <TM1>
F:30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental
F:555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 3,2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWYG 24
Db 346 GLFGAIGFIENGWEGMIDGWYG 368

RESULT 14

hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C:Accession: A93231; A04051
R:Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, E.; Fiers
Nature 286, 771-776, 1980
A:Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strain A/Verhoeven, M.; Fang, R.; Min Jou, W.; Devos, R.; Huybrecock, D.; Saman, E.; Fiers
A:Reference number: A93231; MUID:80254693
A:Accession: A93231
A:Molecule type: genomic RNA
A:Residues: 1-566 <VER>
A:Cross-references: GB:J02090; NID:q324131; PIDN:AAA43178.1; PID:q324132
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>
 F:555,562/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match

55.1%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 3.2e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
 |||||
 DB 346 GLFGAIGFIENGWEGMIDGMYG 368

RESULT 15

HMIYHM

hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)

N:Contains: hemagglutinin HA1; hemagglutinin HA2

C:Species: influenza A virus

C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000

C/Accession: A94441; A04051

R:Steigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.

in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, F

A:Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of C

A:Reference number: A94441

A:Accession: A94441

A:Molecule type: genomic RNA

A:Residues: 1-566 <SDE>

C:Superfamily: influenza virus hemagglutinin

C:Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>

F:346-566/Product: hemagglutinin HA2 #status predicted <HA2>

F:555,562/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match

55.1%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 3.2e-06;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDGMYG 24
 |||||
 DB 346 GLFGAIGFIENGWEGMIDGMYG 368

Search completed: July 1, 2002, 06:28:12
 Job time: 714 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:31:35 ; Search time 25.37 Seconds

(without alignments)
67.153 Million cell updates/sec

Title: US-09-461-684-5

Perfect score: 243
Sequence: 1 CGTGAAGACIENGEGMDID.....KKKKKKKKKKKKKKKKKK 44

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134	55.1	550	HEMA_IADH1	P12582 influenza a
2	134	55.1	550	HEMA_IADH2	P12583 influenza a
3	134	55.1	550	HEMA_IADH3	P12584 influenza a
4	134	55.1	550	HEMA_IADH4	P12585 influenza a
5	134	55.1	550	HEMA_IADH5	P12586 influenza a
6	134	55.1	550	HEMA_IADH6	P12587 influenza a
7	134	55.1	550	HEMA_IADH7	P12588 influenza a
8	134	55.1	550	HEMA_IADH8	P12589 influenza a
9	134	55.1	550	HEMA_IADH9	P12590 influenza a
10	134	55.1	550	HEMA_IADH10	P12591 influenza a
11	134	55.1	550	HEMA_IADH11	P12592 influenza a
12	134	55.1	550	HEMA_IADH12	P12593 influenza a
13	134	55.1	550	HEMA_IADH13	P12594 influenza a
14	134	55.1	550	HEMA_IADH14	P12595 influenza a
15	134	55.1	550	HEMA_IADH15	P12596 influenza a
16	134	55.1	550	HEMA_IADH16	P12597 influenza a
17	134	55.1	550	HEMA_IADH17	P12598 influenza a
18	134	55.1	550	HEMA_IADH18	P12599 influenza a
19	134	55.1	550	HEMA_IADH19	P12600 influenza a
20	134	55.1	550	HEMA_IADH20	P12601 influenza a
21	134	55.1	550	HEMA_IADH21	P12602 influenza a
22	134	55.1	550	HEMA_IADH22	P12603 influenza a
23	134	55.1	550	HEMA_IADH23	P12604 influenza a
24	134	55.1	550	HEMA_IADH24	P12605 influenza a
25	134	55.1	550	HEMA_IADH25	P12606 influenza a
26	134	55.1	550	HEMA_IADH26	P12607 influenza a
27	134	55.1	550	HEMA_IADH27	P12608 influenza a
28	134	55.1	550	HEMA_IADH28	P12609 influenza a
29	134	55.1	550	HEMA_IADH29	P12610 influenza a
30	134	55.1	550	HEMA_IADH30	P12611 influenza a
31	134	55.1	550	HEMA_IADH31	P12612 influenza a
32	134	55.1	550	HEMA_IADH32	P12613 influenza a
33	134	55.1	550	HEMA_IADH33	P12614 influenza a

34	131	53.9	565	HEMA_IADH34	P16996 influenza a
35	131	53.9	565	HEMA_IADH35	P16997 influenza a
36	131	53.9	565	HEMA_IADH36	P16998 influenza a
37	131	53.9	565	HEMA_IADH37	P16999 influenza a
38	131	53.9	565	HEMA_IADH38	P17000 influenza a
39	131	53.9	565	HEMA_IADH39	P17001 influenza a
40	131	53.9	565	HEMA_IADH40	P17002 influenza a
41	131	53.9	565	HEMA_IADH41	P17003 influenza a
42	131	53.9	565	HEMA_IADH42	P17004 influenza a
43	131	53.9	565	HEMA_IADH43	P17005 influenza a
44	131	53.9	565	HEMA_IADH44	P17006 influenza a
45	131	53.9	565	HEMA_IADH45	P17007 influenza a

ALIGNMENTS

RESULT 1
HEMA_IADH1 STANDARD; PRT; 550 AA.
AC P12582; Q84021; Q84022;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor (Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAZ chain) (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11357;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87265458; PubMed=2440178;
RX Kida H., Kawakura Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks";
RT virology 159:109-119(1987).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC -!- CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: M16737; AAA43143.1; -
CC PIR: A27813; HMI177.
CC InterPro: IPR001364; Hemagglutn.
CC Pfam: PF00509; Hemagglutinin; 1.
CC ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein.
KW NON_TER
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 8 22 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 22 38 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 36 38 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 550 AA: 61705 MW: 767ACEF716FC969A CRG64;
SO

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGMYG 24
 Db 330 GLFGAIGFIENGEGMIDGMYG 352

RESULT 2

HEMA_IADH2 STANDARD; PRT; 550 AA.

AC P12583; 084011; 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAZ chain] (Fragment).
 GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/8/80).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A virus.

OC Influenza A virus.
 OX NCBI_TaxID=11358;

RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=67265458; PubMed=2440178;
 RA Kida H., Kawoka Y., Naeye C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";

RL Virology 159:109-119(1987).

CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL: M16738; AAA43144.1; -

DR PIR: B27813; HMTV80.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; glycoprotein.

KW NON_TER 1

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CARBOHYD 330 550 HEMAGGLUTININ HAZ CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).

FT SEQUENCE 550 AA; 61659 MW; A107023ACC9CC353 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;

Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGEGMIDGMYG 24
 Db 330 GLFGAIGFIENGEGMIDGMYG 352

RESULT 3

HEMA_IADH3

ID HEMA_IADH3 STANDARD; PRT; 550 AA.

AC P12584; 084012; 089793; 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/33/80).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus.
 OX NCBI_TaxID=11359;

RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=67265458; PubMed=2440178;
 RA Kida H., Kawoka Y., Naeye C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";

RL Virology 159:109-119(1987).

CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL: M16739; AAA43145.1; -

DR PIR: C27813; HMTV33.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR ProDom: PD000225; Hemagglutn; 1.

KW Envelope protein; Hemagglutinin; glycoprotein.

KW NON_TER 1

FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.

FT CARBOHYD 330 550 HEMAGGLUTININ HAZ CHAIN.

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).

FT SEQUENCE 550 AA; 61577 MW; 6C30BF67CFCB7DE CRC64;

RESULT 4

HEMA_IADH4

ID HEMA_IADH4 STANDARD; PRT; 550 AA.

AC P12585; 084013; 084014; 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAZ chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/7/82).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

```
CC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus;
OX NCBI_TaxID=11360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16740; AAA43146.1; -.
DR PIR: D27813; HMIY89.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61664 MW; A16B2CF8CBBD9D0 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 5
HEMA_IADH5 STANDARD; PRT; 550 AA.
ID HEMA_IADH5
AC P12586; Q84015; Q84016;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus;
OX NCBI_TaxID=11361;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
```

```
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16741; AAA43147.1; -.
DR PIR: E27813; HMIY21.
DR HSSP: P03437; 5SMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 388 388 YV -> VI (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61856 MW; 48401C867A15B8C CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIENGWEGMIDGWYG 24
DB 330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 6
HEMA_IADH6 STANDARD; PRT; 550 AA.
ID HEMA_IADH6
AC P12587; Q84017;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
  Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus;
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawacka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
  ducks.";
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
  CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
  (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M16743; AAA43148.1; -
DR PIR: F27813; HMTV98.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 8 8 Y -> N (IN PIR DATA BANK).
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFTENGEGMIDGNG 24
Db 330 GLFGAIAGFTENGEGMIDGNG 352

RESULT 7
HEMA_IADH7 STANDARD; PRT; 550 AA.
ID HEMA_IADH7
AC P12568; O84018; O89470;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_Taxid=11363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8725458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RT J. Virol. 159:109-119 (1987).
RL CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M16743; AAA43149.1; -
DR PIR: G27813; HMTV15.

DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61761 MW; 6EF81793281D53EB CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFTENGEGMIDGNG 24
Db 330 GLFGAIAGFTENGEGMIDGNG 352

RESULT 8
HEMA_IADHK STANDARD; PRT; 550 AA.
ID HEMA_IADHK
AC P43257;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_Taxid=11364;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/58 (H3N2) strain emerged.";
RT J. Gen. Virol. 72:2007-2010 (1991).
RL CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D00929; BAA00769.1; -
DR HSSP: P03437; SHMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CARBOHYD 330 550
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 550 AA: 61549 MW: 864639823FE1BA9 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24
Db 330 GLFGAIGFIENGWEGMIDWYG 352

RESULT 9
HEMA_IADHL STANDARD; PRT; 550 AA.
AC P43258;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
CN HA.
OS Influenza A virus (strain A/Duck/Hong Kong/54/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=45412;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged."
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D00931; BAA00771.1; -
DR HSSP: P03437; 3HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR Prodom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 550 AA: 61718 MW: A351C56789E4BE9A CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFIENGWEGMIDWYG 24
Db 330 GLFGAIGFIENGWEGMIDWYG 352

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
HEMA_IAGHK STANDARD; PRT; 550 AA.
AC P43260;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
CN HA.
OS Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=45414;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91341491; PubMed=1875195;
RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT "Molecular evidence for a role of domestic ducks in the introduction
RT of avian H3 influenza viruses to pigs in southern China, where the
RT A/Hong Kong/68 (H3N2) strain emerged."
RL J. Gen. Virol. 72:2007-2010(1991).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D00930; BAA00770.1; -
DR HSSP: P03437; 3HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR Prodom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 550 AA: 61676 MW: 9A1ED094DA28BACD2 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIGFIENGWEGMIDWYG 24
Db 330 GLFGAIGFIENGWEGMIDWYG 352

RESULT 11
HEMA_IADHL

```
ID HEMA_IJAZH2 STANDARD; PRT; 550 AA.
AC P1133; Q84019; Q84020;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11497;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19057; AAA43212.1; -
DR PIR: B29971; HMTV53.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 550 AA; 61437 MW; 1F2A7E758C531CE8 CRC64;
SQ
Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFTENGWEGMIDGWY 24
DB 330 GLFGAIGFTENGWEGMIDGWY 352
RESULT 12
HEMA_IJAZH3 STANDARD; PRT; 550 AA.
AC P1134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
```

```
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88101364; PubMed=3336940;
RA Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
in China."
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMTV52.
DR HSP: P03437; 2HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 550 AA; 61560 MW; 991F658BC0242F2 CRC64;
SQ
Query Match 55.1%; Score 134; DB 1; Length 550;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLFGAIGFTENGWEGMIDGWY 24
DB 330 GLFGAIGFTENGWEGMIDGWY 352
RESULT 13
HEMA_IJAZH3 STANDARD; PRT; 556 AA.
AC P03437; 2HMG.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Aich/2/68).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=150147;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80254693; PubMed=7402351;
RA Verhoeven M., Fang R., Min Jou W., Devos R., Huybrecock D.,
Saman E., Fiers W.;
RT "Antigenic drift between the haemagglutinin of the Hong Kong
```



```
FT STRAND 310 311
FT STRAND 318 320
FT STRAND 323 324
FT STRAND 331 333
FT STRAND 337 337
FT STRAND 347 348
FT TURN 350 350
FT TURN 351 351
FT TURN 352 354
FT STRAND 355 355
FT STRAND 359 359
FT TURN 360 361
FT STRAND 367 372
FT TURN 374 375
FT STRAND 378 382
FT STRAND 383 401
FT STRAND 406 407
FT HELIX 421 474
FT STRAND 475 477
FT STRAND 482 485
FT HELIX 491 498
FT TURN 499 500
FT HELIX 505 515
SQ SEQUENCE 566 AA; 63415 MW; E395659C23CAFECA CRC64;
```

```
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2 GLFGAAGFIENGMEGMIDWYG 24
DB 346 GLFGAAGFIENGMEGMIDWYG 368
```

```
RESULT 14
HEMA_IAD3 STANDARD; PRT; 566 AA.
ID P26134;
AC 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
OS HA.
GN Influenza A virus (strain A/Duck/Alberta/79/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11348;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schell M., Katz J., Kawacka Y., Naeve C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
nonhuman hosts".
RL J. Virol. 66:1129-1138(1992).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M73771; -, NOT_ANNOTATED_CDS.
```

```
DR HSP; P03437; SHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63534 MW; FE19AB6FF9415B89 CRC64;
```

```
Query Match 55.1%; Score 134; DB 1; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 2 GLFGAAGFIENGMEGMIDWYG 24
DB 346 GLFGAAGFIENGMEGMIDWYG 368
```

```
RESULT 15
HEMA_IAD3 STANDARD; PRT; 566 AA.
ID P03442;
AC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].
OS HA.
GN Influenza A virus (strain A/Duck/Ukraine/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11374;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8202542; PubMed=6169439;
RA Pang R., Min Jou W., Huybrecock D., Devos R., Fiers W.;
RA "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
hemagglutinin".
RL Cell 25:315-323(1981).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: Y01087; CAA24271.1; -
DR PIR; A04053; HMIYDU.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
```

FT	CHAIN	17	344	HEMAGGLUTININ HA1 CHAIN.
FT	CHAIN	346	566	HEMAGGLUTININ HA2 CHAIN.
FT	CARBOHYD	24	24	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	54	54	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	97	97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	566 AA:	63530 MW;	E70F87F0AE117BF4 CRC64;

Query Match 55.1%; Score 134; DB 1; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDWTG 24
 |||||
 DB 346 GLFGAIGFIENGWEGMIDWTG 368

Search completed: July 1, 2002, 06:31:35
 Job time: 672 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:30:59 ; Search time 83.98 Seconds
(without alignments)
90.638 Million cell updates/sec

Title: US-09-461-684-5
Perfect score: 243
Sequence: 1 CGFAGAIAGTENGWGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	55.1	566	12 Q9DHG0	Q9dhg0 Influenza a
2	134	55.1	566	12 Q67125	Q67125 Influenza a
3	134	55.1	566	12 Q67126	Q67126 Influenza a
4	134	55.1	566	12 Q67132	Q67132 Influenza a
5	134	55.1	566	12 Q98052	Q98052 Influenza a
6	134	55.1	566	12 Q81M7	Q81m7 Influenza a
7	134	55.1	566	12 Q810M5	Q810m5 Influenza a
8	133	54.3	501	12 Q9DXK3	Q9dxk3 Influenza a
9	132	54.3	550	12 Q82499	Q82499 Influenza a
10	132	54.3	550	12 Q82753	Q82753 Influenza v
11	132	54.3	550	12 Q82498	Q82498 Influenza a
12	132	54.3	566	12 Q82496	Q82496 Influenza a
13	132	54.3	571	12 Q03909	Q03909 Influenza a
14	131	53.9	109	12 Q67050	Q67050 Influenza a
15	131	53.9	109	12 Q67051	Q67051 Influenza a
16	131	53.9	109	12 Q67052	Q67052 Influenza a

17	131	53.9	109	12 Q67053	Q67053 Influenza a
18	131	53.9	362	12 Q82513	Q82513 Influenza a
19	131	53.9	362	12 Q82517	Q82517 Influenza a
20	131	53.9	362	12 Q84174	Q84174 Influenza a
21	131	53.9	362	12 Q9QKD3	Q9qkd3 Influenza a
22	131	53.9	362	12 Q9QKD2	Q9qkd2 Influenza a
23	131	53.9	362	12 Q9QKD1	Q9qkd1 Influenza a
24	131	53.9	365	12 Q9DL25	Q9dl25 Influenza a
25	131	53.9	367	12 Q9DL22	Q9dl22 Influenza a
26	131	53.9	368	12 Q9DL29	Q9dl29 Influenza a
27	131	53.9	369	12 Q9DL26	Q9dl26 Influenza a
28	131	53.9	369	12 Q9DL06	Q9dl06 Influenza a
29	131	53.9	369	12 P87689	P87689 Influenza a
30	131	53.9	371	12 Q9DL24	Q9dl24 Influenza a
31	131	53.9	371	12 P87685	P87685 Influenza a
32	131	53.9	373	12 Q9DL20	Q9dl20 Influenza a
33	131	53.9	374	12 Q9DL21	Q9dl21 Influenza a
34	131	53.9	375	12 Q9DL27	Q9dl27 Influenza a
35	131	53.9	375	12 Q9DL05	Q9dl05 Influenza a
36	131	53.9	376	12 Q9DL30	Q9dl30 Influenza a
37	131	53.9	376	12 Q9DL04	Q9dl04 Influenza a
38	131	53.9	377	12 Q9E7P6	Q9e7p6 Influenza a
39	131	53.9	382	12 Q9DL03	Q9dl03 Influenza a
40	131	53.9	408	12 Q9E7P5	Q9e7p5 Influenza a
41	131	53.9	409	12 Q9Q0L5	Q9q0l5 Influenza a
42	131	53.9	416	12 Q9QJ04	Q9qj04 Influenza a
43	131	53.9	429	12 Q9Q0L4	Q9q0l4 Influenza a
44	131	53.9	438	12 Q9Q0L3	Q9q0l3 Influenza a
45	131	53.9	467	12 Q9E312	Q9e312 Influenza a

ALIGNMENTS

RESULT 1
ID Q9DHG0 PRELIMINARY; PRT; 566 AA.
AC Q9DHG0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HAEMAGGLUTININ PRECURSOR.
OS Influenza A virus H3N2.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
OX NCBI_taxid=41857;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLONE 7A (H3N2);
RA Molsin M.A., Morris S.J., Smith H., Sweet C.;
RT "Influenza virus-induced apoptosis: a dual role for viral
RT neuraminidase".
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HAEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HAEMAGGLUTININ FAMILY.
CC EMBL; AJ289703; CAC18525.1; -.
DR HSSP; P03437; 2V1U.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HAEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT SIGNAL 1
FT SEQUENCE 566 AA; 63356 MW; 0BA601929300F72F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24
 DB 346 GLFGAIGFIENGMEGMDGMYG 368

RESULT 2

O67125 PRELIMINARY; PRT; 566 AA.

AC 067125; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DR HEMAGGLUTININ.

GN HA.

OC Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3911/92;

RL MEDLINE=95146951; PubMed=7844533;

RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL: L31943; AAA64228.1; -

DR HSSP: P03437; 2V10.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR PRODOM: PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

SO SEQUENCE 566 AA; 63456 MW; AE56302A9EB89F CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24

DB 346 GLFGAIGFIENGMEGMDGMYG 368

RESULT 3

O67126 PRELIMINARY; PRT; 566 AA.

AC 067126; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DR HEMAGGLUTININ.

GN HA.

OC Influenza A virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses.

OX NCBI_TaxID=11320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/SEAL/MA/3984/92;

RL MEDLINE=95146951; PubMed=7844533;

RT "The appearance of H3 influenza viruses in seals."

RL J. Gen. Virol. 76:199-203(1995).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL: L32024; AAA64228.1; -

DR HSSP: P03437; 2V10.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR PRODOM: PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

SO SEQUENCE 566 AA; 63441 MW; 590576CB4CEB7D08 CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24

DB 346 GLFGAIGFIENGMEGMDGMYG 368

RESULT 4

O67132 PRELIMINARY; PRT; 566 AA.

AC 067132; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DR HEMAGGLUTININ.

GN HA.

OC Influenza A virus (strain A/Alchi/2/68).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses; Influenza A virus.

OX NCBI_TaxID=150147;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A/Alchi/2/68;

RL Min J.W., Verhoeven M., Fang R.-X., Devos R., Huylenbroeck D.,

RA Fiers W.;

RT "Shift and drift in influenza viruses."

RL (in) Carille M.J., Collins J.F., Moseley B.E. B. (eds.)

RL SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, PP.285-311,

RL Cambridge University Press, New York (1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

DR EMBL: M55059; AAA43239.1; -

DR HSSP: P03437; 1HGE.

DR InterPro: IPR001364; Hemagglutn.

DR Pfam: PF00509; Hemagglutinin; 1.

DR PRINTS: PR00329; HEMAGGLUTN12.

DR PRODOM: PD000225; Hemagglutn; 1.

KW Envelope protein; Glycoprotein; Hemagglutinin.

FT CHAIN 1 344 HEMAGGLUTININ.

FT CHAIN 1 346 NEURAMINIDASE.

SO SEQUENCE 566 AA; 63441 MW; E5D1B97D96FECA CRC64;

Query Match 55.1%; Score 134; DB 12; Length 566;

Best Local Similarity 100.0%; Pred. No. 4e-08;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGMEGMDGMYG 24

DB 346 GLFGAIGFIENGMEGMDGMYG 368

RESULT 5

O68052 PRELIMINARY; PRT; 566 AA.

AC 068052; 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE HEMAGGLUTININ PRECURSOR (FRAGMENT).
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=81053698; PubMed=6253883;
 RT Both G.W., Sleight M.J.;
 RT "Complete nucleotide sequence of the haemagglutinin gene from a human
 RT Influenza virus of the Hong Kong subtype."
 RL Nucleic Acids Res. 8:2561-2575(1980).
 RN [2]
 RP SEQUENCE OF 17-344 FROM N.A.
 RA MEDLINE=81194918; PubMed=6164798;
 RT Sleight M.J., Both G.W., Underwood P.A., Bender V.J.;
 RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza
 RT subtype: Correlation of amino acid changes with alterations in viral
 RT antigenicity."
 RL J. Virol. 37:845-853(1981).
 RN [3]
 RP SEQUENCE OF 17-566 FROM N.A.
 RA MEDLINE=82033276; PubMed=6169843;
 RT Both G.W., Sleight M.J.;
 RT "Conservation and variation in the hemagglutinins of Hong Kong subtype
 RT Influenza viruses during antigenic drift."
 RL J. Virol. 39:845-853(1981).
 RN [4]
 RP CELL RECEPTORS AND FOR INITIATING THE VIRUS TO
 RN SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL; J02135; AAA3189.1; -;
 DR HSSP; P03437; IHGE.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS; PR00329; HEMAGGLUTN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 344 POTENTIAL.
 FT CHAIN 346 566 POTENTIAL.
 SQ SEQUENCE 566 AA; 63414 MW; C447FD45B5B4FCF9 CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 4e-08;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAIGFIENGMEGMDGMYG 24
 DB 346 GLFGAIGFIENGMEGMDGMYG 368
 RESULT 6
 ID Q91MA7 PRELIMINARY; PRT; 566 AA.
 AC Q91MA7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE HEMAGGLUTININ.
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A/HONG KONG/1/68(H3N2);
 RC MEDLINE=21287244; PubMed=11371620;
 RX Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
 Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 DR EMBL; AF348176; AAK51718.1; -;
 SQ SEQUENCE 566 AA; 63387 MW; 01BB0D465BE158E1 CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 4e-08;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAIGFIENGMEGMDGMYG 24
 DB 346 GLFGAIGFIENGMEGMDGMYG 368
 RESULT 7
 ID Q910M5 PRELIMINARY; PRT; 566 AA.
 AC Q910M5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DE HEMAGGLUTININ.
 OS Influenza A virus (A/Hong Kong/1/68(H3N2)).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=108859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A/HONG KONG/1/68(H3N2);
 RC MEDLINE=21287244; PubMed=11371620;
 RX Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
 RT "Pattern of mutation in the genome of influenza A virus on adaptation
 RT to increased virulence in the mouse lung: Identification of functional
 RT themes."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
 DR EMBL; AF348179; AAK51721.1; -;
 DR EMBL; AF348177; AAK51720.1; -;
 DR EMBL; AF348178; AAK51720.1; -;
 SQ SEQUENCE 566 AA; 63530 MW; 7CB9F5BAF1E6E9F4 CRC64;
 Query Match 55.1%; Score 134; DB 12; Length 566;
 Best Local Similarity 100.0%; Pred. No. 4e-08;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLFGAIGFIENGMEGMDGMYG 24
 DB 346 GLFGAIGFIENGMEGMDGMYG 368
 RESULT 8
 ID Q9DXE3 PRELIMINARY; PRT; 301 AA.
 AC Q9DXE3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE HEMAGGLUTININ (FRAGMENT).
 GN HA1.
 OS Influenza A virus (A/Shorebird/Taiwan/31-4/99).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses; Influenza A virus.
 OX NCBI_TaxID=140665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A/SHOREBIRD/TAIWAN/31-4/99;
 RC Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee J.H., Shieh H.K.,
 RT "Identification and subtyping of avian influenza virus by reverse
 RT transcription-polymerase chain reaction."
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

```

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: AF311750; AAG3016.1; -.
DR InterPro: IPR001364; Hemagglutn.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 301
FT SEQUENCE 301 AA; 32701 MW; 62A403758B764D57 CRC64;

QY Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 2.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 250 GIEGALGFIENGMEGMDGMYG 24
|||||
RESULT 9
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of influenza A virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule.";
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
RL CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: J08889; AA18782.1; -.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 328
FT CHAIN 1 328 HA1.
FT CHAIN 330 550 HA2.
FT SEQUENCE 550 AA; 61772 MW; 50BD62B6FE11FD8 CRC64;

QY Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 6.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 330 GIEGALGFIENGMEGMDGMYG 24
|||||
RESULT 11
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of influenza A virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule.";
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);

```

```

AC Q82753;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ (FRAGMENT).
OS Influenza virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Unclassified Orthomyxoviridae.
OX NCBI_TaxID=11305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RX MEDLINE=97300854; Pubmed=9155874;
RA Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RT virus associated with increased virulence for mice.";
RL Arch. Virol. 142:75-86(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82/BS/ML10 (H3N2);
RX MEDLINE=97456249; Pubmed=9311563;
RA Ward A.C.;
RT "Virulence of influenza A virus for mouse lung.";
RL Virus Genes 14:187-194(1997).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR EMBL: J08905; AAC79579.1; -.
DR HSSP: P03437; 2V1U.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Glycoprotein; Hemagglutinin.
FT NON_TER 1 328
FT CHAIN 1 328 HAEMAGGLUTININ HA1.
FT CHAIN 330 550 HAEMAGGLUTININ HA2.
FT SEQUENCE 550 AA; 61745 MW; 692A49DE678AC4BC CRC64;

QY Query Match 54.3%; Score 132; DB 12; Length 550;
Best Local Similarity 95.7%; Pred. No. 6.7e-08;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 330 GIEGALGFIENGMEGMDGMYG 352
|||||
RESULT 11
Q82498 PRELIMINARY; PRT; 550 AA.
AC Q82498;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE HEMAGGLUTININ HA1 AND HA2 (FRAGMENT).
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses.
OX NCBI_TaxID=11320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);
RA Hartley C.A., Ward A.C., Anders E.M.;
RT "Virulence of influenza A virus for mice is associated with loss of
RT oligosaccharide from the hemagglutinin molecule.";
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A/PHILIPPINES/2/82 (H3N2);

```

RX MEDLINE=8185444; PubMed=3356226;
 RA Nakajima S., Takeuchi Y., Nakajima K.;
 RT "Location on the evolutionary tree of influenza H3 haemagglutinin
 RL genes of Japanese strains isolated during the 1985-6 season."
 CC Epidemiol. Infect. 100:301-310(1988).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: U08856; AAA18781.1; -.
 DR HSSP: P03437; 2V1U.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTIN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1 1
 FT CHAIN 1 328 HA1.
 FT CHAIN 330 350 HA2.
 SQ SEQUENCE 550 AA; 61802 MW; 1144131CE5A1F6A CRC64;

Query Match 54.3%; Score 132; DB 12; Length 550;
 Best Local Similarity 95.7%; Pred. No. 6.7e-08;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAAGFIENGEGMIDGWTG 24
 1:|||||
 DB 330 GFGAAGFIENGEGMIDGWTG 352

RESULT 12
 Q82496 PRELIMINARY; PRT; 566 AA.
 AC Q82496;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ.
 GN HA.
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_Taxid=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/SWINE/ANGE-GARDIEN/150/90(H3N2);
 RA MEDLINE=95205091; PubMed=7897358;
 RA Brikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,
 RA Elazhary Y.;
 RT "Recent H3N2 swine influenza virus with haemagglutinin and
 RT nucleoprotein genes similar to 1975 human strains."
 RT J. Gen. Virol. 76:697-703(1995).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: U07146; AAA85781.1; -.
 DR HSSP: P03437; 2V1U.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTIN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Glycoprotein; Hemagglutinin.
 SQ SEQUENCE 566 AA; 65362 MW; 1FB8319A561E2FF CRC64;

Query Match 54.3%; Score 132; DB 12; Length 566;
 Best Local Similarity 95.7%; Pred. No. 6.9e-08;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAAGFIENGEGMIDGWTG 24
 1:|||||
 DB 346 GFGAAGFIENGEGMIDGWTG 368

RESULT 13
 Q03909 PRELIMINARY; PRT; 571 AA.
 ID Q03909;
 AC Q03909;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ PRECURSOR.
 GN HA.
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_Taxid=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Guo Y., Wang M., Kawachi Y., Gorman O.T., Ito T., Webster R.G.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 DR EMBL: M65018; AAA43151.1; -.
 DR HSSP: P03437; 2V1U.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTIN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 349 HA1 CHAIN.
 FT CHAIN 350 571 HA2 CHAIN.
 FT DISULFID 36 487 BY SIMILARITY.
 FT DISULFID 73 298 BY SIMILARITY.
 FT DISULFID 85 97 BY SIMILARITY.
 FT DISULFID 160 494 BY SIMILARITY.
 SQ SEQUENCE 571 AA; 64104 MW; 718DAA0F291CE349 CRC64;

Query Match 54.3%; Score 132; DB 12; Length 571;
 Best Local Similarity 95.7%; Pred. No. 7e-08;
 Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GFGAAGFIENGEGMIDGWTG 24
 1:|||||
 DB 351 GFGAAGFIENGEGMIDGWTG 373

RESULT 14
 Q67050 PRELIMINARY; PRT; 109 AA.
 ID Q67050;
 AC Q67050;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ (FRAGMENT).
 GN HA.
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_Taxid=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/FUKUOKA/C29/85 (H3N2);
 RA MEDLINE=81030832; PubMed=7421990;
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin
 RT genes from h2 and h3 strains elucidates antigenic shift and drift in
 RT human influenza virus."
 RL Nature 287:301-306(1980).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=A/FUKUOKA/C29/85 (H3N2);
 RX MEDLINE=93333219; PubMed=7682624;
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
 RT "A common neutralizing epitope conserved between the hemagglutinins of
 influenza A virus H1 and H2 strains.";
 RL J. Virol. 67:2552-2558(1993).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: D13581; BAA02776.1; -.
 DR HSSP: P03437; IHTV.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTIN12.
 DR ProDom: PD000225; Hemagglutn; 1.
 DR Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT 109
 SQ SEQUENCE 109 AA; 12305 MW; 17EC66753C48672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;
 Best Local Similarity 91.3%; Pred. No. 1.5e-08;
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
 DB 40 GIFGAIAGFIENGWEGMVDGWTG 62

RESULT 15
 ID 067051 PRELIMINARY; PRT; 109 AA.
 AC 067051;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEMAGGLUTININ (FRAGMENT).
 GN HA.
 OS Influenza A virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses.
 OX NCBI_TaxID=11320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/SICHUAN/2/87 (H3N2);
 RX MEDLINE=81030852; PubMed=7421990;
 RA Gething M.-J., Bye J., Skehel J., Waterfield M.;
 RT "Cloning and dna sequence of double-stranded copies of haemagglutinin
 genes from h2 and h3 strains elucidates antigenic shift and drift in
 human influenza virus.";
 RT Nature 287:301-306(1980).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=A/SICHUAN/2/87 (H3N2);
 RX MEDLINE=93333219; PubMed=7682624;
 RA Okuno Y., Isegawa Y., Sasao F., Ueda S.;
 RT "A common neutralizing epitope conserved between the hemagglutinins of
 influenza A virus H1 and H2 strains.";
 RL J. Virol. 67:2552-2558(1993).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 DR EMBL: D13582; BAA02777.1; -.
 DR HSSP: P03437; IHGE.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR PRINTS: PR00329; HEMAGGLUTIN12.

DR ProDom: PD000225; Hemagglutn; 1.
 KW Envelope protein; Glycoprotein; Hemagglutinin.
 FT NON_TER 1
 FT 109
 SQ SEQUENCE 109 AA; 12293 MW; 17EC66752DB8672F CRC64;

Query Match 53.9%; Score 131; DB 12; Length 109;
 Best Local Similarity 91.3%; Pred. No. 1.5e-08;
 Matches 21; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAIGFIENGWEGMIDGWTG 24
 DB 40 GIFGAIAGFIENGWEGMVDGWTG 62

Search completed: July 1, 2002, 06:30:59
 Job time: 691 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:19:25 ; Search time 98.97 Seconds

(without alignments)
49.381 Million cell updates/sec

Title: US-09-461-684-5
Perfect score: 243
Sequence: 1 CGLFGAIGFTENGEGMID.....KKKKKKKKKKKKKKKKKK 44

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	243	100.0	44	21	AA13784
2	192	79.0	40	22	AA151431
3	143	58.8	24	21	AA13782
4	143	58.8	26	13	AA26004
5	143	58.8	26	18	AA134269
6	143	58.8	26	18	AA134269
7	143	58.8	26	18	AA134269
8	134	55.1	25	17	AA198016
9	134	55.1	26	13	AA198005
10	134	55.1	26	18	AA134270
11	134	55.1	26	22	AA170091

12	134	55.1	28	17	AA198024	Fusogenic peptide
13	134	55.1	30	21	AA198680	Core polypeptide f
14	134	55.1	30	22	AA198088	Viral P178/107-11
15	134	55.1	30	22	AA198257	Viral core polypep
16	134	55.1	30	22	AA198634	P178-11ke/D107-1
17	134	55.1	30	22	AA198081	Core polypeptide T
18	134	55.1	175	15	AA198579	Sequence of SHA2.
19	134	55.1	223	12	AA194032	Peptide sequence f
20	134	55.1	347	15	AA198391	Stem region of A2/
21	134	55.1	566	15	AA198390	Full length H3N2 i
22	134	55.1	685	22	AA198783	Modified clostridi
23	132	54.3	29	20	AA198710	Influenza virus an
24	132	54.3	221	14	AA198864	Sequence of the HA
25	132	54.3	221	14	AA198865	Sequence of the HA
26	132	54.3	221	15	AA198020	Influenza haemagl
27	132	54.3	221	15	AA198021	Influenza haemagl
28	132	54.3	306	14	AA198867	Sequence of fusion
29	132	54.3	306	15	AA198014	Immunogenic fragme
30	132	54.3	386	5	AA198015	Sequence of the X-
31	131	53.9	365	8	AA198071	Equine influenza v
32	131	53.9	565	11	AA198493	Equine hemagglutin
33	131	53.9	565	19	AA198494	EIV Fontainebleau
34	131	53.9	565	21	AA198056	Wild type equine i
35	131	53.9	565	21	AA198057	Cold-adapted equin
36	131	53.9	566	19	AA198406	SIV strain H3N2 ha
37	131	53.9	570	8	AA198070	Equine influenza v
38	131	53.9	570	11	AA198490	Equine hemagglutin
39	131	53.9	570	18	AA198169	Influenza A/Beijin
40	131	53.9	570	20	AA198441	Influenza virus A/
41	131	53.9	570	22	AA198491	Influenza virus A/
42	131	53.9	571	18	AA198167	Influenza A/Shando
43	131	53.9	571	18	AA198166	Influenza A/Joahn
44	131	53.9	571	20	AA198445	Influenza virus A/
45	131	53.9	571	20	AA198448	Influenza virus A/

ALIGNMENTS

RESULT 1	AA13784	standard; peptide: 44 AA.
ID	AA13784	
XX	AA13784	
AC	10-NOV-2000	(first entry)
XX		
DT		
DE	Soluble tandem HA/ PK peptide conjugate.	
XX		
KW	pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;	
KW	major histocompatibility complex class I; MHC class I; antigen; tumour;	
KW	prostate; breast; multiple myeloma; HA peptide.	
XX		
OS	Unidentified.	
XX		
PN	WO200035949-A1.	
XX		
PD	22-JUN-2000.	
XX		
PF	14-DEC-1999;	99MO-US29724.
XX		
PR	14-DEC-1998;	98US-0112324.
XX		
PA	(DEND-) DENDREON CORP.	
XX		
PI	Laus R, Hakim I, Vidovic D;	
XX		
DR	WPI, 2000-442365/38.	
XX		
PT	Antigens modified by the covalent addition of a peptide that	
PT	facilitates entry into antigen presenting cells, useful for producing	
PT	compositions for immunizing against tumors and pathogens -	

PS Claim 2; Page 26; 34pp; English.

The present invention relates to compositions of modified soluble protein antigens capable of eliciting an enhanced *in vivo* cytotoxic T cell (CTL) response i.e. a major histocompatibility complex (MHC) class I molecule response. The protein antigen is modified by the covalent addition of a peptide sequence which facilitates entry of the antigen into antigen presenting cells (APCs). The present sequence is one such peptide sequence which can be used to modify the soluble antigens. The present sequence is tandem HA/pK peptide conjugate. The modified antigen composition may be used for immunising against, or treating a tumour e.g. prostate and breast carcinoma or multiple myeloma, or pathogen in mammals.

Query Match	100.0%;	Score 243;	DB 21;	Length 44;
Best Local Similarity	100.0%;	Pred. No. 7.4e-23;		
Matches 44;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	CGLFGAIAGFIENGWEGMIDGWYKXXXXXXXXXXXXXXXXXXXXX	44
Db	1	cglfgaiagfiengwegmidgwygxkxxxxxxxxxxxxxxxxxxxxx	44

RESULT 2

ID AAM51431 standard; peptide; 40 AA.

DT 28-JAN-2002 (first entry)

Peptide K14 Fuso influenza haemagglutinin fusogenic peptide.

KM Gene therapy vector; cell entry; intracellular trafficking
KM gene expression; influenza; haemagglutinin.

OS Synthetic.

PN WO200149324-A2.

PD 12-JUL-2001.

PF 28-DEC-2000; 2000MO-EP13300.

PR 30-DEC-1999; 99US-0475305.

PA (NOVS) NOVARTIS AG.-----

PI Woodle M, Cheng C, Puthupparampil S, Subramanian K, Titmas R

DR WPT; 2001-602251/68

PT Non-naturally occurring gene therapy vector useful for gene therapy
PT comprises an inner shell having a core complex containing a nucleic
PT acid and at least one complex forming reagent -

CC The invention relates to a non-naturally occurring gene therapy vector,
CC comprising an inner shell having a core complex containing a nucleic acid
CC and at least one complex forming reagent. The vectors are stable having
CC an improved outer steric layer that provides enhanced target specificity,
CC in vivo and colloidal stability. The vectors are relatively homogeneous
CC and comprise chemically defined species. The vectors demonstrate improved
CC cell entry and intracellular trafficking, permitting enhanced nucleic
CC acid therapeutic activity such as gene expression. The present sequence
CC is that of the peptide K14 Fusco, comprising a fusogenic peptide derived
CC from Influenza haemagglutinin and peptide K14, useful in the preparation

CC of coated core complexes for the invention.

Sequence 40 AA

Query Match	79.0%;	Score 192;	DB 22;	Length 40;
Best Local Similarity	90.0%;	Pred. No. 1.1e-16;		
Matches 36;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

```
OY      2 GLEGAIAETIENGDEGMIDGWIKKKKKKKKKKKKK 41
        |||||  |||||  |||||  |||||  |||||
DB      1 gLfgalegfienegwldgwygcckkkkkkkkkkkk 40
```

RESULT 3

ID AAB13782 standard; peptide; 24 AA.

AC AAB13782

XX	Soluble peptide antigen HA
DE	

OS . Unidentified.

PN WO200035949-A1

PD 22-JUN-2000

PF 14-DEC-1999; 99WO-US29724.

PR 14-DEC-1998; 98US-0112324

PA (DEND-) DENDREON CORP.

PI Laus R, Hakim I, Vidovic D

DR WP1; 2000-442365/38

PT Antigens modified by the covalent addition of a peptide that
PT facilitates entry into antigen presenting cells, useful for producing
PT compositions for immunizing against tumors and pathogens -

CC The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced *in vivo* cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC response. The protein antigen is modified by the covalent addition of a
CC peptide sequence which facilitates entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is peptide HA. The modified antigen composition may be used for
CC immunising against, or treating a tumour e.g. prostate and breast
CC carcinoma or multiple myeloma, or pathogen in mammals.

Query Match	58.8%;	Score 143;	DB 21;	Length 24;
Best Local Similarity	100.0%;	Pred. No. 6.4e-11;		
Matches	24;	Conservative	0;	Mismatches
			0;	Indels
				0;
				Gaps
				0;

OY 1 CGLFGAIAGFIENGWEGMIDGWYC 24
|||
Db 1 cglfgaiagflengwegmidgwys 24

RESULT * 4
AAR26004
ID AAR26004 standard; peptide; 26 AA.
XX
AC AAR26004;
XX
DT 26-JAN-1993 (first entry)
XX
DE Influenza fusion peptide #1.
XX
KM Haemagglutinin; cysteine; liposome.
XX
OS Synthetic.
XX
PN EP497997-A.
XX
PD 12-AUG-1992.
XX
PF 02-FEB-1991; 91EP-0101414.
XX
PR 02-FEB-1991; 91EP-0101414.
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
PI Gluck R, Herrmann P, Klein P;
XX
DR WPI; 1992-270078/33.
XX
XX Drug-contg. phospholipid bilayer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS
XX
PS Disclosure; Fig 2; 13pp; English.
XX
CC The sequences given in AAR26004-19 are influenza fusion peptides,
CC derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cysteine residues at one
CC end of these peptides has been found useful for the fusion activity,
CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.
XX
SQ Sequence 26 AA;
XX
Query Match 58.8%; Score 143; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CGLFGAIGFIENGEGMIDGWY 24
DB 3 cglfgaigfiengewgmldgwyg 26
XX
RESULT 5
AAR34269
ID AAR34269 standard; peptide; 26 AA.
XX
AC AAR34269;
XX
DT 14-MAY-1998 (first entry)
XX
DE Synthetic lipid vesicle fusion peptide 1.
XX
KM virosome; vesicle; cationic lipid bilayer; viral fusion peptide;
KM drug delivery system; membrane; gene therapy; diagnosis; treatment;
KM cancer; leukaemia; viral infection.
XX
OS Synthetic.
XX

PN W09741834-A1.
XX
XX
PD 13-NOV-1997.
XX
XX
PF 04-MAY-1997; 97WO-EP02268.
XX
PR 08-MAY-1996; 96EP-0107282.
XX
XX
PA (NIKA-) NIKA HEALTH PROD LTD.
XX
XX
PI Glueck R, Klein P, Maeltl ER;
XX
DR WPI; 1997-558673/51.
XX
XX
XX Vesicle with cationic lipid bilayer that includes viral fusion
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukaemia and viral infections
XX
XX
PS Disclosure; Page 10; 52pp; English.
XX
XX
CC Peptides AAR34269-W34284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment)
CC of cancer, leukaemia and viral infections in humans or animals. Genetic
CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.
XX
SQ Sequence 26 AA;
XX
Query Match 58.8%; Score 143; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CGLFGAIGFIENGEGMIDGWY 24
DB 3 cglfgaigfiengewgmldgwyg 26
XX
RESULT 6
AAB70090
ID AAB70090 standard; peptide; 26 AA.
XX
XX
AC AAB70090;
XX
DT 14-MAY-2001 (first entry)
XX
DE Cationic virosome crosslinker #1 for polypeptide binding.
XX
KM virosome; vesicle; cytostatic; gene therapy; lipid bilayer vesicle;
KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;
KM crosslinker.
XX
XX
OS unidentified.
XX
PN NZ504444-A.
XX
PD 24-NOV-2000.
XX

XX AA98010-R98041 represent fusogenic peptides derived from spike
CC glycoproteins of enveloped viruses. These fusogenic peptides act as
CC endosome membrane disruption promoting components (EMDPCs). These
CC sequences were used in a transfer moiety which makes up part of the
CC multifunctional molecular complex of the invention. The multifunctional
CC molecular complex is used for the transfer of a nucleic acid composition
CC to a target cell, and comprises the nucleic acid composition and a
CC transfer moiety comprising one or more cationic polyamine components
CC bound to the nucleic acid composition. There are several different
CC cationic polyamines that can be used in the complex, one of which
CC contains one of these sequences linked via a bridging group to the main
CC body of the cationic polyamine. The products can be used for delivering
CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth
CC regulators, insecticides, miticides, rodenticides, fungicides,
CC parasiticides or nematocides. They can also be used for immunising an
CC individual against a pathogen or for treating an autoimmune disease or
CC hyperproliferating disease. The complexes provide for a high level of
CC transfection and expression of the nucleic acid molecules in target
CC cells.

XX Sequence 25 AA:

Query Match 55.1%; Score 134; DB 17; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGWEGMDGWYG 24
Db 1 glfgaagfiengwgmldgwyg 23

RESULT 9

AA26005
ID AA26005 standard; peptide; 26 AA.

XX AAR26005;

XX 26-JAN-1993 (first entry)

XX Influenza fusion peptide #2.

XX Haemagglutinin; cysteine; liposome.

XX Synthetic.

XX EP497997-A.

XX 12-AUG-1992.

XX 02-FEB-1991; 91EP-0101414.

XX 02-FEB-1991; 91EP-0101414.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Gluck R, Herrmann P, Klein P;

XX WPI: 1992-270078/33.

XX Drug-contg. phospholipid bi-layer vesicle with cell-specific
PT markers on membrane - where markers have at least 90 per cent
PT biological activity, used as pharmaceuticals against e.g. cancer
PT and AIDS

XX Disclosure; Fig 2; 13pp; English.

XX The sequences given in AA26004-19 are influenza fusion peptides,
CC derived from the influenza virus haemagglutinin gene which were
CC used for the preparation of synthetic membrane vesicles. The
CC arrangement of at least one, pref. three cysteine residues at one
CC end of these peptides has been found useful for the fusion activity,

CC for the fusion of the liposome to the target cell membrane. The
CC liposomes produced using these peptides can contain at least one
CC active drug and can be used to target cells infected with cancer
CC or AIDS.

XX Sequence 26 AA:

Query Match 55.1%; Score 134; DB 13; Length 26;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLFGAAGFIENGWEGMDGWYG 24
Db 1 glfgaagfiengwgmldgwyg 23

RESULT 10

AAW34270
ID AAW34270 standard; peptide; 26 AA.

XX AAW34270;

XX 14-MAY-1998 (first entry)

XX Synthetic lipid vesicle fusion peptide 2.

XX Virosome; vesicle; cationic lipid bilayer; viral fusion peptide;

XX drug delivery system; membrane; gene therapy; diagnosis; treatment;

XX cancer; leukaemia; viral infection.

XX Synthetic.

XX WO9741834-A1.

XX 13-NOV-1997.

XX 04-MAY-1997; 97WO-EP02268.

XX 08-MAY-1996; 96EP-0107282.

XX (NIKA-) NIKA HEALTH PROD LTD.

XX Glueck R, Klein P, Waelte ER;

XX WPI: 1997-558673/51.

XX Vesicle with cationic lipid bilayer that includes viral fusion
PT peptide - used for delivery of genetic material to cells, especially
PT for gene therapy of cancer, leukaemia and viral infections

XX Disclosure; Page 10; 52pp; English.

XX Peptides AAW34269-W34284 represent novel lipid vesicles with positively
CC charged lipid bilayer membranes composed of a cationic and/or
CC polycationic lipid and at least one natural or synthetic viral fusion
CC peptide integrated in, or covalently linked to, the membrane. Such
CC peptides are used as drug delivery systems, preferably for
CC (non-)specific delivery of genetic material to target cells or tissues,
CC particularly for diagnosis, treatment (especially antisense treatment)
CC of cancer, leukaemia and viral infections in humans or animals. Genetic
CC material is delivered, without infection, to resting or proliferating
CC cells, in vitro or in vivo. When the genetic material is an antisense
CC molecule, it is targeted to mRNA encoding a (proto)oncogene. The
CC continuous lipid layer does not leak. The peptides do not need to fuse
CC with, or destabilise, plasma membranes in order to enter the cytoplasm,
CC since the fusion peptide ensures cell penetration by endocytosis (after
CC which fusion of the vesicle and endosomal membrane occurs). The genetic
CC material thus has a greater chance of reaching the nucleus before it is
CC degraded or expelled. Transfer of the material is 1000-20000 times more
CC efficient than when using liposomes or conventional virosomes, so
CC smaller doses can be used, avoiding possible toxicity associated with the
CC genetic material.

XX Sequence 26 AA;

Query Match 55.1%; Score 134; DB 18; Length 26;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLFGAIAGFIEGMEGMDGWG 24
Db 1 glfgaiaagiengwegmidgwyg 23

RESULT 11
AAB70091 standard; peptide: 26 AA.

XX AAB70091;

DT 14-MAY-2001 (first entry)

DE Cationic virosome crosslinker #2 for polypeptide binding.

KW Cationic virosome; cytosolic; gene therapy; lipid bilayer vesicle;

KM viral haemagglutinin; drug delivery; cancer; leukaemia; viral infection;

XX crosslinker.

CS Unidentified.

XX NZ504444-A.

PN 24-NOV-2000.

PE 10-MAY-2000; 2000NZ-0504444.

XX 10-MAY-2000; 2000NZ-0332666.

PA (NIKA-) NIKA HEALTH PROD LTD.

PI Waiiti ER, Gluck R, Klein P;

DR WPI: 2001-233042/24.

PT Lipid bilayer vesicle, useful for delivering drugs to target cells such

PS as cancer, leukemic, or virally infected cells, comprises viral

XX glycoproteins in positively charged membrane

CC Disclosure; Page 14; 41pp; English.

CC The present sequence is a crosslinker which is capable of linking to the

CC surface of a novel cationic virosome and is capable of binding

CC polypeptides. The cationic virosome is a lipid bilayer vesicle comprising

CC a membrane with a net positive charge and contains 5-30 weight % based on

CC total lipids, 1,3-di-oleoyloxy-2-(6-carboxy-spermyl)-propyl-amine

CC (DOSPER) together with other lipids and at least one active fusogenic

CC peptide. The fusogenic peptide is a viral haemagglutinin that causes the

CC vesicle to be internalised by target cells through phagocytosis or

CC endocytosis. The virosome is useful for delivering a desired drug or

CC substance, preferably a nucleic acid, to target cells (resting or

CC proliferating mammalian cells) such as cancer cells, leukemic cells or

CC virally infected cells in vitro, in diagnostic or medical applications

CC and for the manufacture of medicament for carrying out non-infectious

CC gene therapy.

Db 1 glfgaiaagiengwegmidgwyg 23

RESULT 12
AAR98024 standard; peptide: 28 AA.

XX AAR98024;

DT 12-NOV-1996 (first entry)

DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.

KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;

KM endosome membrane disruption promoting component; cationic polyamine;

KW multifunctional molecular complex; foodstuff; herbicide; insecticide;

KM plant growth regulator; miticide; rodenticide; fungicide; parasiticide;

KW nematocide; immunisation; pathogen; therapy; autoimmune disease;

KM hyperproliferating disease.

XX Synthetic.

XX WO9610038-A1.

PN 04-APR-1996.

PD 28-SEP-1995; 95WO-US12502.

XX 28-SEP-1994; 94US-0314060.

PA (APOL-) APOLLON INC.

PI Boutin RH;

DR WPI: 1996-200887/20.

PT New complexes for nucleic acid transfer to target cells - comprising

XX a nucleic acid compsn. and a cationic poly:amine with an endosome

XX membrane disruption component.

PS Disclosure; Page 30; 138pp; English.

CC AAR98010-R98041 represent fusogenic peptides derived from spike

CC glycoproteins of enveloped viruses. These fusogenic peptides act as

CC endosome membrane disruption promoting components (EMDPCs). These

CC sequences were used in a transfer moiety which makes up part of the

CC multifunctional molecular complex of the invention. The multifunctional

CC molecular complex is used for the transfer of a nucleic acid composition

CC to a target cell, and comprises the nucleic acid composition and a

CC transfer moiety comprising one or more cationic polyamine components

CC bound to the nucleic acid composition. There are several different

CC cationic polyamines that can be used in the complex, one of which

CC contains one of these sequences linked via a bridging group to the main

CC body of the cationic polyamine. The products can be used for delivering

CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth

CC regulators, insecticides, miticides, rodenticides, fungicides,

CC parasiticides or nematocides. They can also be used for immunising an

CC individual against a pathogen or for treating an autoimmune disease or

CC hyperproliferating disease. The complexes provide for a high level of

Db 1 glfgaiaagiengwegmidgwyg 23

RESULT 12
AAR98024 standard; peptide: 28 AA.

XX AAR98024;

DT 12-NOV-1996 (first entry)

DE Fusogenic peptide derived from HA-2 X:31 spike glycoprotein.

KW Fusogenic peptide; spike glycoprotein; enveloped virus; vaccine; EMDPC;

KM endosome membrane disruption promoting component; cationic polyamine;

KW multifunctional molecular complex; foodstuff; herbicide; insecticide;

KM plant growth regulator; miticide; rodenticide; fungicide; parasiticide;

KW nematocide; immunisation; pathogen; therapy; autoimmune disease;

KM hyperproliferating disease.

XX Synthetic.

XX WO9610038-A1.

PN 04-APR-1996.

PD 28-SEP-1995; 95WO-US12502.

XX 28-SEP-1994; 94US-0314060.

PA (APOL-) APOLLON INC.

PI Boutin RH;

DR WPI: 1996-200887/20.

PT New complexes for nucleic acid transfer to target cells - comprising

XX a nucleic acid compsn. and a cationic poly:amine with an endosome

XX membrane disruption component.

PS Disclosure; Page 30; 138pp; English.

CC AAR98010-R98041 represent fusogenic peptides derived from spike

CC glycoproteins of enveloped viruses. These fusogenic peptides act as

CC endosome membrane disruption promoting components (EMDPCs). These

CC sequences were used in a transfer moiety which makes up part of the

CC multifunctional molecular complex of the invention. The multifunctional

CC molecular complex is used for the transfer of a nucleic acid composition

CC to a target cell, and comprises the nucleic acid composition and a

CC transfer moiety comprising one or more cationic polyamine components

CC bound to the nucleic acid composition. There are several different

CC cationic polyamines that can be used in the complex, one of which

CC contains one of these sequences linked via a bridging group to the main

CC body of the cationic polyamine. The products can be used for delivering

CC therapeutic agents, vaccines, foodstuffs, herbicides, plant growth

CC regulators, insecticides, miticides, rodenticides, fungicides,

CC parasiticides or nematocides. They can also be used for immunising an

CC individual against a pathogen or for treating an autoimmune disease or

CC hyperproliferating disease. The complexes provide for a high level of

RESULT 13
 AAY89680
 ID AAY89680 standard; peptide; 30 AA.
 AC AAY89680;
 XX
 DT 23-MAY-2000 (first entry)
 DE Core polypeptide fragment T No. 1242.
 XX
 KM Retrovirus; hybrid polypeptide; enhancer; gp41; envelope protein; HIV-1;
 KM HIV-2; SIV; pharmacokinetic; half-life; growth factor; cytokine; viral;
 KM anti-fusogenic; differentiation factor; interleukin; interferon;
 KW colony stimulating factor; hormone; angiogenic factor.
 XX
 OS Unidentified.
 XX
 PN WO9959615-A1.
 XX
 PD 25-NOV-1999.
 XX
 PF 20-MAY-1999; 99WO-US11219.
 XX
 PR 20-MAY-1998; 98US-0082279.
 XX
 PA (TRIM-) TRIMERIS INC.
 XX
 PI Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;
 XX
 DR WPI; 2000-136792/12.
 XX
 PT A new hybrid polypeptide with enhanced pharmacokinetic properties
 PT comprises enhancer sequence -
 XX
 PS Disclosure; Page 42; 124pp; English.
 XX
 CC The invention relates to hybrid polypeptides comprising enhancer peptide
 CC sequence linked to core polypeptides. The enhancer polypeptides are
 CC derived from various retroviral envelope (gp41) protein sequences,
 CC especially from HIV-1, HIV-2 and SIV. The enhancer peptides enhance the
 CC pharmacokinetic properties such as increasing the half-life of any core
 CC polypeptide that they are linked to. The core polypeptides are any
 CC polypeptide that may be introduced into a living system and that can
 CC function as a pharmacologically useful peptide for the treatment or
 CC prevention of a disease. The core polypeptides are bioactive peptides
 CC selected from a growth factor, cytokine, differentiation factor,
 CC interleukin, interferon, colony stimulating factor, hormone or
 CC angiogenic factor. The peptides of the invention can be used for
 CC inhibiting viral infection and can be used in anti-viral and
 CC anti-fusogenic treatments. Sequences AAY88651-Y9005 represent core
 CC polypeptide fragments that can be used in the invention. Some sequences
 CC among those indicated also comprise enhancer fragments at terminal ends
 CC and form hybrid polypeptides.
 CC
 XX
 SQ Sequence 30 AA;
 XX
 QY Query Match 55.1%; Score 134; DB 21; Length 30;
 XX Best Local Similarity 100.0%; Pred. No. 1e-09;
 XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 glfgaiafienwgmgidgwyg 23
 XX
 RESULT 14
 ABB01088
 ID ABB01088 standard; Peptide; 30 AA.
 AC ABB01088;
 XX

DT 03-JAN-2002 (first entry)
 XX
 DE Viral DP178/107-like region peptide T1242.
 XX
 KM Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;
 KM virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;
 KW infection.
 XX
 OS Viridae.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT /note= "N-terminal is substituted by Ac"
 FT Modified-site 30
 FT /note= "C-terminal amide"
 XX
 XX WO200164013-A2.
 XX
 PN 07-SEP-2001.
 XX
 PD 07-FEB-2001; 2001WO-US03988.
 XX
 PF 29-FEB-2000; 2000US-0515965.
 XX
 PR (TRIM-) TRIMERIS INC.
 XX
 PA Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;
 XX
 PI WPI; 2001-514829/56.
 XX
 DR Heptad repeat region peptide analogs useful for inhibiting virus/cells
 XX fusion, useful for treating HIV and Respiratory Syncytial Virus
 PT infection -
 PT
 XX
 PS Disclosure; Page 55; 587pp; English.
 XX
 CC The invention relates to isolated analogues of the heptad repeat region
 CC peptides DP178 and DP107. DP178 and DP107 correspond to amino acids
 CC 638-673 (heptad repeat region HR2) and 558-595 (heptad repeat region
 CC HR1) respectively, of HIV-1LAI transmembrane protein gp41. The HR1
 CC and HR2 regions of proteins interact non-covalently with each other
 CC and/or with peptides derived from them. This interaction is required for
 CC normal infectivity of viruses such as RSV and HIV. The heptad
 CC repeat region peptide analogues may be used to inhibit respiratory
 CC syncytial virus (RSV) infection in a cell. They may also be used to
 CC inhibit HIV infection. The present sequence is a peptide provided in
 CC the specification.
 CC
 XX
 SQ Sequence 30 AA;
 XX
 QY Query Match 55.1%; Score 134; DB 22; Length 30;
 XX Best Local Similarity 100.0%; Pred. No. 1e-09;
 XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 glfgaiafienwgmgidgwyg 23
 XX
 RESULT 15
 ABB02537
 ID ABB02537 standard; Peptide; 30 AA.
 AC ABB02537;
 XX
 DT 03-JAN-2002 (first entry)
 DE Viral core polypeptide, SEQ ID NO: 1064.
 XX
 DE Human immunodeficiency virus; HIV; respiratory syncytial virus; RSV;
 KW virucide; heptad repeat region; transmembrane protein; gp41; HR1; HR2;
 KW infection.

XX Viridae.
 OS
 XX
 PN W0200164013-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 07-FEB-2001; 2001MO-US03988.
 XX
 PR 29-FEB-2000; 2000US-0515965.
 XX
 PA (TRIM-) TRIMERIS INC.
 XX
 PI Antczak JB, Delmedico MK, Erickson JB, Lambert DM, Sista P;
 XX
 DR WPI; 2001-514829/56.
 XX
 PT Heptad repeat region peptide analogs useful for inhibiting virus/cells
 PT fusion, useful for treating HIV and Respiratory Syncytial Virus
 PT infection -
 XX
 PS Disclosure; Page 448; 587pp; English.
 XX
 CC The invention relates to isolated analogues of the heptad repeat region
 CC peptides DP178 and DP107. DP178 and DP107 correspond to amino acids
 CC 638-673 (heptad repeat region HR2) and 558-595 (heptad repeat region
 CC HR1) respectively of HIV-1/IIA1 transmembrane protein gp41. The HR1
 CC and HR2 regions of proteins interact non-covalently with each other
 CC and/or with peptides derived from them. This interaction is required for
 CC normal infectivity of viruses such as RSV and HIV. The heptad
 CC repeat region peptide analogues may be used to inhibit respiratory
 CC syncytial virus (RSV) infection in a cell. They may also be used to
 CC inhibit HIV infection. The present sequence is a peptide provided in
 CC the specification.
 XX
 SQ Sequence 30 AA;

Query Match 55.1%; Score 134; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1e-09;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GTFGAIAGFIENGWEGMTDGYG 24
 ||||||||||||||||||||
 Db 1 gtfgaiaagiengwegmtdgyg 23

Search completed: July 1, 2002, 06:15:25
 Job time: 512 sec

! FINDPATTERNS on geneseqp: * allowing 0 mismatches
! 1 C(R,K){20} July 1, 2002 07:07 ..

1 AA45801 ck: 9500 len: 39 ! Aa45801 One chain of a bombesin dimer. 6/1
C(R,K){20}
C(R,K){20}
1: CXXXXXXXXXXXXXXXXX XGGGQ

1 AAB13780 ck: 7317 len: 21 ! Aab13780 Soluble peptide antigen PK. 11/200
C(R,K){20}
C(R,K){20}
1: CXXXXXXXXXXXXXXXXX

1 AAU18238 ck: 5509 len: 58 ! Aau18238 Novel human DNA-binding protein #8
C(R,K){20}
C(R,K){20}
35: KFYFV CXXXXXXXXXXXXXXXXX KKK

1 AA003766 ck: 8808 len: 81 ! Aa003766 Human polypeptide SEQ ID NO 17658.
C(R,K){20}
C(R,K){20}
48: LTTTA CXXXXXXXXXXXXXXXXX KKKK

1 AA011210 ck: 863 len: 70 ! Aa011210 Human polypeptide SEQ ID NO 25102.
C(R,K){20}
C(R,K){20}
30: IDLCL CXXXXXXXXXXXXXXXXX KKKK

Databases searched:
EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 5
Total length: 114,001,827
Total sequences: 766,495
CPU time: 07:03.10

Seq #6

11AA_SEQUENCE 1.0
ID AAW45801 standard; peptide; 39 AA.
XX
AC AAW45801;
XX
DT 25-JUN-1998 (first entry)
XX
DE One chain of a bombesin dimer.
XX
KW Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;
KW alpha-MSH-AMT; bombesin; dimer; bivalent agonist; disulphide bond;
KW G-protein coupled receptor.
XX
OS Synthetic.
XX
FH Key 1 Location/Qualifiers
FT Cross-links 1 /note= "This residue is disulphide bonded to the
FT corresponding Cys residue of an identical chain
FT to form a dimer."
FT Modified-site 28 /note= "Epsilon-aminohexanoic acid"
XX
PN WO9803632-A1.
XX
PD 29-JAN-1998.
XX
PE 23-JUL-1997; 97WO-US12911.
XX
PR 24-JUL-1996; 96US-0686934.
XX
PA (UYVA) UNIV YALE.
XX
PI Carriers MD, Lerner MR;
XX
DR WPI; 1998-120757/11.
XX
DT 1998-120757/11.
XX
PT Bivalent agonist of G-protein coupled receptors containing two
PT ligand domains - bonded to molecular backbone, for treatment of
PT hypotension, promotion of skin tanning etc., also for delivering
PT drugs and gene therapy vectors to selected cells
XX
PS Claim 41; Page 48; 71pp; English.
XX
CC This sequence represents one of two identical chains disulphide bonded
CC to form a bombesin dimer. The invention relates to bivalent agonists,
CC with affinity for at least 1 G-protein coupled receptor (GPCR). The
CC bivalent agonists comprise: (a) two ligand domains (LD), individually
CC agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)
CC a molecular backbone (MB) covalently bound to LD. The bivalent agonists
CC are useful in human or veterinary medicine as carriers for drugs or gene
CC therapy vectors, allowing these to be endocytosed by GPCR-expressing
CC cells. They can also be used e.g. to treat hypertension (angiotensin-
CC based LD); to increase levels of luteinising hormone (LH), using LD
CC derived from LH-releasing hormone, or to promote skin tanning (LD based
CC on alpha-melanocyte-stimulating hormone, MSH). The bivalent agonists
CC are administered orally, by injection or topically. Typical doses for
CC skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or
CC the bivalent agonists are administered topically in a composition
CC containing 0.001-10 (especially 1) mM, where both LD are agonists, the
CC bivalent agonist has a synergistically higher activity than two
CC individual agonist ligands, and where at least 1 is an antagonist the
CC effect is stimulatory. The bivalent agonists are active at lower
CC concentrations than known agonists so should avoid toxicity problems.
XX
SQ Sequence 39 AA;
AAW45801 Length: 39 July 1, 2002 14:50 Type: P Check: 9500 ..
1 CKRKKKKKK KXGGGQRRG NQWAVGHM
11AA_SEQUENCE 1.0
ID AAB13780 standard; peptide; 21 AA.

XX
AC AAB13780;
XX
DT 10-NOV-2000 (first entry)
XX
DE Soluble peptide antigen pK.
XX
KW pK peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW prostate; breast; multiple myeloma.
XX
OS Unidentified.
XX
PN WO200035949-A1.
XX
PD 22-JUN-2000.
XX
PE 14-DEC-1999; 99WO-US29724.
XX
PR 14-DEC-1999; 98US-0112324.
XX
PA (DEND-) DENDREON CORP.
XX
PI Laus R, Hakim I, Vidovic D;
XX
DR WPI; 2000-442365/38.
XX
PT Antigens modified by the covalent addition of a peptide that
PT facilitates entry into antigen presenting cells, useful for producing
PT compositions for immunising against tumors and pathogens -
XX
PS Claim 2; Page 26; 34pp; English.
XX
CC The present invention relates to compositions of modified soluble protein
CC antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC response i.e. a major histocompatibility complex (MHC) class I molecule
CC response. The protein antigen is modified by the covalent addition of a
CC peptide sequence which facilitate entry of the antigen into antigen
CC presenting cells (APCs). The present sequence is one such peptide
CC sequence which can be used to modify the soluble antigens. The present
CC sequence is peptide pK. The modified antigen composition may be used for
CC immunising against, or treating a tumour e.g. prostate and breast
CC carcinoma or multiple myeloma, or pathogen in mammals.
XX
SQ Sequence 21 AA;
AAB13780 Length: 21 July 1, 2002 14:50 Type: P Check: 7317 ..
1 CKRKKKKKK KKKKKKKKK K
11AA_SEQUENCE 1.0
ID AAU18238 standard; Protein; 58 AA.
XX
AC AAU18238;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human DNA-binding protein #85.
XX
KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein;
KW DNA organisation; gene transcription; malignant disease;
KW autoimmune disorder; rheumatic disease; genetic abnormality;
KW infectious disease; neurological disorder; gene therapy;
KW immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW cytostatic.
XX
OS Homo sapiens.
XX
PN WO200155162-A1.
XX
PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-0S01305.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0216135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250330.
PR 05-DEC-2000; 2000US-0251093.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX MPI: 2001-465557/50.
DR N-PSDB; AAS29114.
XX

PT Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
PS
PS Claim 11: SEQ ID NO 223; 561pp; English.
XX
CC The present invention relates to the isolation of novel DNA-binding
CC proteins, and cDNA (AA29030-AA29157) and genomic sequences encoding
CC for these proteins. DNA-binding proteins such as histones, chromo
CC (chromatin organisation modifier) domain proteins, and Y-box binding
CC proteins may contribute to diseases resulting from aberrant DNA
CC organisation and/or gene transcription. The sequences of the invention
CC are useful in screening assays to identify antagonists and/or agonists
CC that may enhance or block activities mediated by DNA-binding proteins.
CC Blockers of DNA-binding proteins may be useful in treating disorders
CC such as malignant diseases (e.g. cancer), autoimmune disorders
CC (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
CC arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
CC disease). The polynucleotide sequences of the invention may also be
CC used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
CC proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 58 AA:
AAU18238 Length: 58 July 1, 2002 14:50 Type: P Check: 5509 ..
1 TYLECEHNSL VNSKCLTVL SRCISVCLNK FYVCKKKKK KKKKKKKKK
51 KKKKKKKK
!!AA_SEQUENCE 1.0
ID AA003766 standard; Protein: 81 AA.
XX
AC AA003766;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 17658.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PE 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR N-PSDB; AAI83697.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PS Claim 20: SEQ ID NO 17658; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation and which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 81 AA:
AA003766 Length: 81 July 1, 2002 14:50 Type: P Check: 8808 ..
1 GLNQTOLRKI LAYSSITHG XIIVLPYP NITILNLTIV ILLTTACKK
51 KKKKKKKKK KKKKKKKKK KKKKKGGG A
!!AA_SEQUENCE 1.0
ID AA011210 standard; Protein: 70 AA.
XX
AC AA011210;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 25102.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PE 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR N-PSDB; AAI91141.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
PS Claim 20: SEQ ID NO 25102; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation and which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 70 AA;

AA011210 Length: 70 July 1, 2002 14:50 Type: P Check: 863 ..

1 YIHRTVFM CMNGLKDNV DKYTDLCG KKKKKKKK KKKKKKKK

51 KKKKKKKK KKKPGGGA

```

!!SEQUENCE LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!      1 C(R,K){20}
                                July 1, 2002 14:38 ..

GENESEQP1906:AAW45801  ck: 9500  len: 39   finds: 1   ! Aaw45801 One chain of a bomb
GENESEQP2000:AAI13780  ck: 7317  len: 21   finds: 1   ! Aai13780 Soluble peptide anti
GENESEQP2001:AAU18238  ck: 5509  len: 58   finds: 1   ! Aau18238 Novel human DNA-bind
GENESEQP2001:AAO03766  ck: 8808  len: 81   finds: 1   ! Aao03766 Human polypeptide SE
GENESEQP2001:AAO11210  ck: 863   len: 70   finds: 1   ! Aao11210 Human polypeptide SE

\\End of list

Databases searched:
  EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds:      5
Total length:    114,001,827
Total sequences: 766,495
CPU time:        09:53.67
  
```

! FINDPATTERNS on pir:* allowing 0 mismatches
! 1 C(R,K){20}

July 1, 2002 06:14 ..

Databases searched:

NR/ Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds: 0
Total length: 96,089,334
Total sequences: 283,138
CPU time: 04:30.92

! FINDPATTERNS on swp:* allowing 0 mismatches
! 1 C(R,K)(20)

July 1, 2002 06:15 ..

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002
SPRMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 0
Total length: 211,714,479
Total sequences: 667,446
CPU time: 10:17.12

! FINDPATTERNS on pir: * allowing 0 mismatches

! 1 (R,K){20}

July 1, 2002 06:48 ..

T49173 ck: 4143 len: 517 1 hypothetical protein T20N10.250 - Arabidops

1 (R,K){20}

444: FERVG KKKKKKKKKKKKKKKKKKK KKKIR

(K){20}

445: ERYGK KKKKKKKKKKKKKKKKKKK KKKIRL

(K){20}

446: RVGKK KKKKKKKKKKKKKKKKKKK KIRLN

(K){20}

447: VGKKK KKKKKKKKKKKKKKKKKKK IRLNF

S58321 ck: 1384 len: 126 1 probable membrane protein YOR309c - yeast

1 (R,K){20}

53: RKRRT RKRRRRRRRRRRRRRRRRR KRSPR

(R,K){20}

54: KRRTT KRRRRRRRRRRRRRRRRRR RSPRR

(R,K){20}

55: RRTTR KRRRRRRRRRRRRRRRRRR SPRRR

T46395 ck: 7330 len: 380 1 hypothetical protein DKFZp434I1120.1 - huma

1 (R,K){20}

355: NLLQ KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

356: LLLQ KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

357: LLLQ KKKKKKKKKKKKKKKKKKK KKKK

(K){20}

358: LQKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

359: QKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

360: KKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

361: KKKK KKKKKKKKKKKKKKKKKKK

I52523 ck: 8048 len: 215 1 nucleoporin p52 homolog - rat (fragment)

1 (R,K){20}

35: CEPLK KKKKKKKKKKKKKKKKKKK KKTGD

(K){20}

36: EPLEK KKKKKKKKKKKKKKKKKKK KTGDN

(K){20}

37: FLEK KKKKKKKKKKKKKKKKKKK TGDNA

Databases searched:

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002
Total finds: 17
Total length: 96,089,334
Total sequences: 283,138
CPU time: 09:02.03

!!SEQUENCE_LIST 1.0
! FINDPATTERNS on pir:* allowing 0 mismatches

1 (R,K)(20)

July 1, 2002 14:57 ..

PIR2:T49173	ck: 4143	len: 517	finds: 4	! hypothetical protein T20N10.25
PIR2:S58321	ck: 1384	len: 126	finds: 3	! probable membrane protein YOR3
PIR2:T46395	ck: 7330	len: 380	finds: 7	! hypothetical protein DKFZp434t
PIR2:I52523	ck: 8048	len: 215	finds: 3	! nucleoporin p62 homolog - rat

\\End of list

Databases searched:

NBRF, Release 71.0, Released on 15Feb2002, Formatted on 19Feb2002

Total finds: 17
Total length: 96,089,334
Total sequences: 283,138
CPU time: 08:41.42

!!AA_SEQUENCE 1.0
 PI:T49173 - hypothetical protein T20N10.250 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
 C:Accession: T49173
 R:D'Angelio, M.; Verzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;
 Rudd, S.; Lemke, K.; Mayer, R.F.X.; Queirer, F.; Salanoubat, M.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 225017
 A:Accession: T49173
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-517 <DAN>
 A:Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
 A:Experimental source: cultivar Columbia; BAC clone T20N10
 C:Genetics:
 A:Gene: ATSP:T20N10.250
 A:Map position: 3
 A:Introns: 312/3; 359/3; 444/3
 C:Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

T49173 Length: 517 July 1, 2002 15:06 Type: P Check: 4143 ..

1 MDLFSLPNE LLYHILSLS TKKALTSVL SKRWNLFAF VPYLEFDDSV
 51 FLHPEERKAE KEGILQSEMD FVDRVLDLHG DGLIKTFSLK CKTGVSDDHV
 101 DMRICVTLAR GVSDDLDFID FRDLVSLPHE VGVSTLVVL RVGSESDLYW
 151 WQKFLCLPWL KTVLDSCWL CIGOFQILL ACPALBELDM TWTWKDSNV
 201 TVSSSLKEL TIDLHGCSV VNKLSFDA PSLYFYCYD SLADYDPQVN
 251 LKNIVEAQIN LLTQAOAQIE VRALNEMLV ADDVEPGLN AMKLITGLRN
 301 VQGLYISPT LEVLSRCEG MPVFNKLVV SIMSDNRNM QAMVLLRNC
 351 PLETLITIG LLHATDKCG DVCCDISRDY KDHSLTSCV KKLQIYEFEG
 401 TIRELEMIKH FLKIFPCLKE MDIYAHENSH TLEKDPITFE RVGRKKKKKK
 451 KKKKKKKKK KKKKKKIRLN FKPYNKTEGF LKRLADKICF IPQCLEFLDV
 501 DSSLGELALL AMDSRPS

!!AA_SEQUENCE 1.0
 PI:S58321 - probable membrane protein YOR309C - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein 06105
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
 C:Accession: S58321; S67215; S71989
 R:Pearson, B.M.; Hernandez, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: S58318
 A:Accession: S58321
 A:Molecule type: DNA
 A:Residues: 1-126 <PEA>
 A:Cross-references: EMBL:X90565; NID:9940836; PID:9940840
 R:Pearson, B.M.; Hernandez, Y.; Kalogeropoulos, A.; Schweizer, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67213
 A:Accession: S67215
 A:Molecule type: DNA
 A:Residues: 1-126 <PEW>
 A:Cross-references: EMBL:Z75217; NID:91420680; PID:91420681;
 MIPS:YOR309C
 A:Experimental source: strain S288C
 R:Pearson, B.M.; Hernandez, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;
 Schweizer, M.
 yeast 12, 1021-1031, 1996
 A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast
 chromosome XV reveals regions of similarity to chromosomes I and XIII.
 A:Reference number: S71986; MUID:97051589

A:Accession: S71989
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-126 <PEF>
 A:Cross-references: EMBL:X90565; NID:9940836; PID:CAA62164.1; PID:9940840
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August
 1995
 C:Genetics:
 A:Map position: 15R
 C:Keywords: transmembrane protein
 F:3-19/Domain: transmembrane #status predicted <TM1>
 F:107-123/Domain: transmembrane #status predicted <TM2>

S58321 Length: 126 July 1, 2002 15:06 Type: P Check: 1384 ..

1 MQLIPQRL LILNPLMMK RKKRKKRKR RERTMMKIP RILKIRRR
 51 RTRRRRRKR RRRRRKKRR RKRSPKKR RRRNKDAFYI LITSPSRSL
 101 LFGFRKFSII IOCLTFSEH ILFHNL

!!AA_SEQUENCE 1.0
 PI:T46395 - hypothetical protein DKFp434I1120.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46395
 R:Ottewaelder, B.; Obermayer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: 223031
 A:Accession: T46395
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-380 <AAA>
 A:Cross-references: EMBL:AL137556
 A:Experimental source: adult testis; clone DKFp434I1120
 C:Genetics:
 A:Note: DKFp434I1120.1

T46395 Length: 380 July 1, 2002 15:06 Type: P Check: 7330 ..

1 MGSTDSKLNK RKAIVQLTK TOPVEATDDA FWDQFMDTA TSVQDFVALV
 51 PAATIRAVRE ESPSNLALIC YKAVEKLVOG AESGCHSEKE KQVLNCSRL
 101 LTRVLPYTFE DDPWKGFEWS TYGAGRGCG EEDDEHARPL AELLALAIAD
 151 LLFCDPTVQ SHRRSTVDSA EDVHSLDCE YIMEAGVGA HSPQPNYIHD
 201 MNRRELKLL LQCFSEAMVL PPAPSGSN PMWQFCSTE NRHALPLFTS
 251 LLNTVCAYDP VGIGIPYNH LFSYREPLV EEAQVILVT LHDSSASSAS
 301 PTVDGTTGT AMDADPGP ENLFVNYLSR IHREDFOFI LKGIARLISN
 351 LLQKRRKKK KKKKKKKKK KKKKKKKKK

!!AA_SEQUENCE 1.0
 FI:15253 - nucleoporin p62 homolog - rat (fragment)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
 C:Accession: 15253
 R:Mang, Z.O.; Akmal, K.M.; Kim, K.H.
 Biol. Reprod. 51, 1022-1030, 1994
 A:Title: An unusual nucleoporin-related messenger ribonucleic acid is present
 in the germ cells of rat testis.
 A:Reference number: 152523; MUID:95151924
 A:Accession: 152523
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-215 <RES>
 A:Cross-references: GB:S75997; NID:9913245; PIDN:AA833384.1; PID:9913246
 A:Experimental source: testis

I52523 Length: 215 July 1, 2002 15:06 Type: P Check: 8048 ..
1 SGGRAATSCD EDCILSSLPF SLGSPVQDC EFLEKRRKK KKKKKKKKK
51 KKKKKTGDN AKSVSRQYSL KYTKLEHEAE QAKVELDFIL SQKLEEDLL
101 SPLEESVKBE SGTIYLQHAD EEREKTYKLA ENIDAQLKRM AODLKDIIEH
151 LNMAGGPADT SDPLQOICKI LNAHMSLQW VDQSSALLOR RVEEASRYCE
201 SRRKEQERSL RIAFD

! FINDPATTERNS on swp:* allowing 0 mismatches

! 1 (R,K){20}

July 1, 2002 06:38 ..

Q12444 ck: 1384 len: 126 ! Q12444 saccharomyces cerevisiae (baker's ye

1 (R,K){20}

53: KKRRT RKRREKRRRRRRRRRRRRR KRSPR

54: KRRTT RKRREKRRRRRRRRRRRRR RSPRK

55: RRTTR RKRREKRRRRRRRRRRRRR SPRKR

Q9P529 ck: 291 len: 128 ! Q9P529 neurospora crassa. hypothetical 15.2

1 (R,K){20}

71: KRRNQ KRRKKKKKKKKKKKKKKKK KKKKK

72: RKNQK KRRKKKKKKKKKKKKKKKK KKKKK

73: KNQKK KRRKKKKKKKKKKKKKKKK KKKKK

74: NQKKK KRRKKKKKKKKKKKKKKKK KKKKK

75: QKKKK KRRKKKKKKKKKKKKKKKK KKKKK

76: KKKKK KRRKKKKKKKKKKKKKKKK KKKKK

77: KKKKK KRRKKKKKKKKKKKKKKKK KKKKK

78: KKKKK KRRKKKKKKKKKKKKKKKK KKKKK

79: KKKKK KRRKKKKKKKKKKKKKKKK KKKKK

80: KKKKK KRRKKKKKKKKKKKKKKKK KKKKK

81: KKKKK KRRKKKKKKKKKKKKKKKK KKEOE

82: KKKKK KRRKKKKKKKKKKKKKKKK KEDES

83: KKKKK KRRKKKKKKKKKKKKKKKK EQESR

Q9HC48 ck: 7602 len: 667 ! Q9HC48 homo sapiens (human). ctcl tumor ant

(R,K){20}

648: GDXTD RRRKKKKKKKKKKKKKKKK

Q9H6Q7 ck: 3351 len: 720 ! Q9H6Q7 homo sapiens (human). cdna: flj21979

(R,K){20}

692: IVSIS KRRKKKKKKKKKKKKKKKK KKKKK

1

693: VSISK KRRKKKKKKKKKKKKKKKK KKKKK

694: SISKK KRRKKKKKKKKKKKKKKKK KKKKK

695: ISKKK KRRKKKKKKKKKKKKKKKK KKKKK

696: SKKKK KRRKKKKKKKKKKKKKKKK KKKKK

697: KKKKK KRRKKKKKKKKKKKKKKKK KKKK

698: KKKKK KRRKKKKKKKKKKKKKKKK KKK

699: KKKKK KRRKKKKKKKKKKKKKKKK KK

700: KKKKK KRRKKKKKKKKKKKKKKKK K

701: KKKKK KRRKKKKKKKKKKKKKKKK

Q9H5V6 ck: 379 len: 168 ! Q9H5V6 homo sapiens (human). cdna: flj22

(R,K){20}

140: VREWE KRRKKKKKKKKKKKKKKKK KKKKK

141: REWEK KRRKKKKKKKKKKKKKKKK KKKKK

142: EWEKK KRRKKKKKKKKKKKKKKKK KKKKK

143: WEKKK KRRKKKKKKKKKKKKKKKK KKKKK

144: EKKKK KRRKKKKKKKKKKKKKKKK KKKKK

145: KKKKK KRRKKKKKKKKKKKKKKKK KKKK

146: KKKKK KRRKKKKKKKKKKKKKKKK KKK

147: KKKKK KRRKKKKKKKKKKKKKKKK KK

148: KKKKK KRRKKKKKKKKKKKKKKKK K

149: KKKKK KRRKKKKKKKKKKKKKKKK

Q9NT34 ck: 7330 len: 380 ! Q9NT34 homo sapiens (human). hypotheticala

(R,K){20}

355: NLLIQ KRRKKKKKKKKKKKKKKKK KKKKK

356: LLLQK KRRKKKKKKKKKKKKKKKK KKKKK

1

357: LLOK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
358: LOKK KKKKKKKKKKKKKKKKK KKK
 (K){20}
359: OKKK KKKKKKKKKKKKKKKKK KKK
 (K){20}
360: KKKK KKKKKKKKKKKKKKKKK K
 (K){20}
361: KKKK KKKKKKKKKKKKKKKKK

Q95LV6 ck: 7515 len: 531 ! Q95lv6 macaca fascicularis (crab eating mac
 (R,K){20}
 (K){20}
502: YKGS KKKKKKKKKKKKKKKKK KKKK
 (K){20}
503: KGSK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
504: GNSK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
505: NSKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
506: SKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
507: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
508: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
509: KKKK KKKKKKKKKKKKKKKKK KKK
 (K){20}
510: KKKK KKKKKKKKKKKKKKKKK KKK
 (K){20}
511: KKKK KKKKKKKKKKKKKKKKK K
 (K){20}
512: KKKK KKKKKKKKKKKKKKKKK

Q9LXR2 ck: 4143 len: 517 ! Q9Lxr2 arabidopsis thaliana (mouse-ear cress
 (R,K){20}
 (K){20}
444: FEVVG KKKKKKKKKKKKKKKKK KKKIR
 (K){20}
445: EVVGK KKKKKKKKKKKKKKKKK KKITL
 (K){20}
446: RVGKK KKKKKKKKKKKKKKKKK KIRLN
 (K){20}
447: VGKKK KKKKKKKKKKKKKKKKK IRLNF
 (K){20}

Q9LGZ9 ck: 6094 len: 260 ! Q9Lgz9 arabidopsis thaliana (mouse-ear cress
 (R,K){20}
 (K){20}
6: MDPCI KKKKKKKKKKKKKKKKK KKKKK

7: DRCIR KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
8: RCIRK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
9: CIRKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
10: IRKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
11: RKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
12: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
13: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
14: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
15: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
18: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
27: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
28: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
29: KKKKK KKKKKKKKKKKKKKKKK KKKKK
 (K){20}
30: KKKKK KKKKKKKKKKKKKKKKK KKKKK

31: KKKK (K){20} KKKK
32: KKKK (K){20} KKKK
33: KKKK (K){20} KKKK
34: KKKK (K){20} KKKK
35: KKKK (K){20} KKKK
36: KKKK (K){20} KKKK
37: KKKK (K){20} KKKK
38: KKKK (K){20} KKKK
39: KKKK (K){20} KKKK
40: KKKK (K){20} KKKK
41: KKKK (K){20} KKKK
42: KKKK (K){20} KKKK
43: KKKK (K){20} KKKK
44: KKKK (K){20} KKKK
45: KKKK (K){20} KKKK
46: KKKK (K){20} KKKK
47: KKKK (K){20} KKKK
48: KKKK (K){20} KKKK
49: KKKK (K){20} KKKK
50: KKKK (K){20} KKKK
51: KKKK (K){20} KKKK
52: KKKK (K){20} KKKK
53: KKKK (K){20} KKKK
54: KKKK (K){20} KKKK

55: KKKK KKKKKKKKKKKKKKKKKKK KKKK
56: KKKK (K){20} KKKK
57: KKKK (K){20} KKKK
58: KKKK (K){20} KKKK
59: KKKK (K){20} KKKK
60: KKKK (K){20} KKKK
61: KKKK (K){20} KKKK
62: KKKK (K){20} KKKK
63: KKKK (K){20} KKKK
64: KKKK (K){20} KKKK
65: KKKK (K){20} KKKK
66: KKKK (K){20} KKKK
67: KKKK (K){20} KKKK
68: KKKK (K){20} KKKK
69: KKKK (K){20} KKKK
70: KKKK (K){20} KKKK
71: KKKK (K){20} KKKK
72: KKKK (K){20} KKKK
73: KKKK (K){20} KKKK
74: KKKK (K){20} KKKK
75: KKKK (K){20} KKKK
76: KKKK (K){20} KKKK
77: KKKK (K){20} KKKK
78: KKKK (K){20} KKKK
79: KKKK (K){20} KKKK

80: KKKK (K){20} KKKK
81: KKKK (K){20} KKKK
82: KKKK (K){20} KKKK
83: KKKK (K){20} KKKK
84: KKKK (K){20} KKKK
85: KKKK (K){20} KKKK
86: KKKK (K){20} KKKK
87: KKKK (K){20} KKKK
88: KKKK (K){20} KKKK
89: KKKK (K){20} KKKK
90: KKKK (K){20} KKKK
91: KKKK (K){20} KKKK
92: KKKK (K){20} KKKK
93: KKKK (K){20} KKKK
94: KKKK (K){20} KKKK
95: KKKK (K){20} KKKK
96: KKKK (K){20} KKKK
97: KKKK (K){20} KKKK
98: KKKK (K){20} KKKK
99: KKKK (K){20} KKKK
100: KKKK (K){20} KKKK
101: KKKK (K){20} KKKK
102: KKKK (K){20} KKKK
103: KKKK (K){20} KKKK

104: KKKK (K){20} KKKK
105: KKKK (K){20} KKKK
106: KKKK (K){20} KKKK
107: KKKK (K){20} KKKK
108: KKKK (K){20} KKKK
109: KKKK (K){20} KKKK
110: KKKK (K){20} KKKK
111: KKKK (K){20} KKKK
112: KKKK (K){20} KKKK
113: KKKK (K){20} KKKK
114: KKKK (K){20} KKKK
115: KKKK (K){20} KKKK
116: KKKK (K){20} KKKK
117: KKKK (K){20} KKKK
118: KKKK (K){20} KKKK
119: KKKK (K){20} KKKK
120: KKKK (K){20} KKKK
121: KKKK (K){20} KKKK
122: KKKK (K){20} KKKK
123: KKKK (K){20} KKKK
124: KKKK (K){20} KKKK
125: KKKK (K){20} KKKK
126: KKKK (K){20} KKKK
127: KKKK (K){20} KKKK

128: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
129: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
130: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
131: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
132: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
133: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
134: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
135: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
136: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
137: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
138: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
139: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
140: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
141: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
142: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
143: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
144: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
145: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
146: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
147: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
148: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
149: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
150: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
151: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
152: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}

153: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
154: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
155: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
156: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
157: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
158: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
159: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
160: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
161: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
162: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
163: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
164: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
165: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
166: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
167: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
168: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
169: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
170: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
171: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
172: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
173: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
174: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
175: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}
176: KKKK KKKKKKKKKKKKKKKKK KKKK
 (K){20}

177: kkkk (K){20} kkkk
178: kkkk (K){20} kkkk
179: kkkk (K){20} kkkk
180: kkkk (K){20} kkkk
181: kkkk (K){20} kkkk
182: kkkk (K){20} kkkk
183: kkkk (K){20} kkkk
184: kkkk (K){20} kkkk
185: kkkk (K){20} kkkk
186: kkkk (K){20} kkkk
187: kkkk (K){20} kkkk
188: kkkk (K){20} kkkk
189: kkkk (K){20} kkkk
190: kkkk (K){20} kkkk
191: kkkk (K){20} kkkk
192: kkkk (K){20} kkkk
193: kkkk (K){20} kkkk
194: kkkk (K){20} kkkk
195: kkkk (K){20} kkkk
196: kkkk (K){20} kkkk
197: kkkk (K){20} kkkk
198: kkkk (K){20} kkkk
199: kkkk (K){20} kkkk
200: kkkk (K){20} kkkk

201: kkkk kkkkkkkkkkkkkkkkk kkkk
202: kkkk (K){20} kkkk
203: kkkk (K){20} kkkk
204: kkkk (K){20} kkkk
205: kkkk (K){20} kkkk
206: kkkk (K){20} kkkk
207: kkkk (K){20} kkkk
208: kkkk (K){20} kkkk
209: kkkk (K){20} kkkk
210: kkkk (K){20} kkkk
211: kkkk (K){20} kkkk
212: kkkk (K){20} kkkk
213: kkkk (K){20} kkkk
214: kkkk (K){20} kkkk
215: kkkk (K){20} kkkk
216: kkkk (K){20} kkkk
217: kkkk (K){20} kkkk
218: kkkk (K){20} kkkk
219: kkkk (K){20} kkkk
220: kkkk (K){20} kkkk
221: kkkk (K){20} kkkk
222: kkkk (K){20} kkkk
223: kkkk (K){20} kkkk
224: kkkk (K){20} kkkk
225: kkkk (K){20} kkkk

(K)(20)
226: KKKKK KKKKKKKKKKKKKKKKK KNRHH

(K)(20)
227: KKKKK KKKKKKKKKKKKKKKKK NRHHI

035807 ck: 7510 len: 129 1 035807 rattus norvegicus (rat). microvascul

1
(R,K)(20)

85: VLLAS KKKKKKKKKKKKKKKKK KIKWE

(K)(20)
86: LLASK KKKKKKKKKKKKKKKKK IKWEG

064075 ck: 8048 len: 215 1 064075 rattus sp. nucleoporin p62 homolog P

1
(R,K)(20)

35: CEPLK KKKKKKKKKKKKKKKKK KKTGD

(K)(20)
36: EPLK KKKKKKKKKKKKKKKKK KTGDN

(K)(20)
37: FLEKK KKKKKKKKKKKKKKKKK TGDNA

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002
SPTREMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286

Total length: 211,714,479

Total sequences: 667,446

CPU time: 18:32.89

```
!!SEQUENCE LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches
!      1 (R,K){20}
```

July 1, 2002 15:28 ..

SP_FUN:Q12444	ck: 1384	len: 126	finds: 3	Q12444	saccharomyces cerevisiae
SP_FUN:Q9P529	ck: 291	len: 128	finds: 13	Q9P529	neurospora crassa. hypc
SP_HUM:Q9HC48	ck: 7602	len: 667	finds: 1	Q9HC48	homo sapiens (human). c
SP_HUM:Q9H6Q7	ck: 3351	len: 720	finds: 10	Q9H6Q7	homo sapiens (human). c
SP_HUM:Q9H5V6	ck: 379	len: 168	finds: 10	Q9H5V6	homo sapiens (human). c
SP_HUM:Q9NT34	ck: 7330	len: 380	finds: 7	Q9NT34	homo sapiens (human). h
SP_OM:Q95LV6	ck: 7515	len: 531	finds: 11	Q95LV6	macaca fascicularis (cr
SP_PL:Q9LXR2	ck: 4143	len: 517	finds: 4	Q9LXR2	arabidopsis thaliana (t
SP_PL:Q9LG29	ck: 6094	len: 260	finds: 222	Q9LG29	arabidopsis thaliana (m
SP_RO:Q35807	ck: 7510	len: 129	finds: 2	Q35807	rattus norvegicus (rat)
SP_RO:Q64075	ck: 8048	len: 215	finds: 3	Q64075	rattus sp. nucleoporin

\\End of list

Databases searched:

SWISS-PROT, Release 40.1, Released on 11Feb2002, Formatted on 19Feb2002
 SPTREMBL, Release 19.0, Released on 14Dec2001, Formatted on 19Feb2002

Total finds: 286
 Total length: 211,714,479
 Total sequences: 667,446
 CPU time: 21:25.76

```

11AA_SEQUENCE 1.0
ID Q12444 PRELIMINARY; PRT; 126 AA.
AC Q12444;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORF YOR309C.
GN YOR309C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearson B.M., Hernandez Y., Kalogeropoulos A., Schweizer M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIRS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=FY1679;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X90565; CAAG629.1; -.
DR EMBL; X90565; CAAG629.1; -.
DR SGD; S0005836; YOR309C.
SQ
SEQUENCE 126 AA; 16294 MW; 46E1F4C664802C8 CRC64;

Q12444 Length: 126 July 1, 2002 15:52 Type: P Check: 1384 ..

1 M0MLP0RL LILNPLIMK RKRKRKRKR RERETMKIP RILKLRKR
51 RRRRRRRR KRRRRRRR RRRRRRRR KRRNKARFY LIISSPSRL
101 LFGFRKPSIT IQCLTFSPH ILFHNH

11AA_SEQUENCE 1.0
ID Q9P529 PRELIMINARY; PRT; 128 AA.
AC Q9P529;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOHETICAL 15.2 KDA PROTEIN.
GN B24H17.160.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algin V., Hohelsel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356815; CAB92636.2; -.
KW Hypothetical protein.
SQ
SEQUENCE 128 AA; 15157 MW; 8C7C65C3DFB70765 CRC64;

Q9P529 Length: 128 July 1, 2002 15:52 Type: P Check: 291 ..

1 MAIISGLHH KNINRAPGH SVYSKSSSYD FQOYDATQHY LPSOGFAIAP
51 DHLTGSKDGC LSTHDKRKNQ KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKEQESRTYF QOHFGADGIC PTPEMHTR

11AA_SEQUENCE 1.0
ID Q9HC48 PRELIMINARY; PRT; 667 AA.

Q9HC48 Length: 667 July 1, 2002 15:52 Type: P Check: 7602 ..

1 EHENVLFREND CIVRINDGL RNRREDAQH MEFQAMRTPI IMFHYPPAN
51 KEQYEQLSOS EKNYYSSRF SPDSQYIDNR SVNSAGLHY QAPRLNHP
101 EQIDSHSRLP HSAHPGKPP SAPASAPQNV FSTVSSGYN TKIKGRINI
151 QLKKGTEGLE FSTSRDVTI GGSAPIYKN ILRGAIOQ GLKAGDRLI
201 EVNGVDIVK SOEYVSLIR STKMEGYSL LVFRQEDAFH PRELNAPBQ
251 M0IPKREKAE DEDIVLTPDG TRELFTEVP LNDGSGAGLG VSKGNSKE
301 NHADLGFYK SIINGGAASK DGRLRVNDQL IAVNGESILG KINODAMETL
351 RRSSTEGNK RGMIOIVAR RISKCNELKS PGSPPGPELP IETALDRRR
401 RISHSLYSGI EGIDESPSRN AALSTINGES GKIQLSFTVN MPDDTVIIE
451 DDRLPVLPFH LSDQSSSSSH DDVGFTVADA GTWAKAASID SADCISLSPV
501 DPVAPFQREG FGRQIADETK LNTYDDCKAG SPSRDVGPSTL GLKSSSLTS
551 LQTVAAEYTL NGDIPHRPR PRIIRGRGN ESFFAAIDKS YKPAVDDDD
601 EGMTELEEDT EESSRGRES VSTASDPSTH SLEROMNGNQ EKGDKTRRK
651 KKKKKKKKK KKKKKKK

11AA_SEQUENCE 1.0
ID Q9H6Q7 PRELIMINARY; PRT; 720 AA.
AC Q9H6Q7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA: FLJ21979 FIS. CLONE HEP06065 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shidohara T.,

```

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK025632; BAB15196.1;
 FT NOVELTER 720 AA; 84029 MW; A86586FEAA953DOB CRC64;
 SO SEQUENCE 720 AA; 84029 MW; A86586FEAA953DOB CRC64;

Q9H607 Length: 720 July 1, 2002 15:52 Type: P Check: 3351

1 MLTEQVEQYT KEKKNITCIT EDLKNELQDN KGASTLSQOT HMKIQSTLDT
 51 LKEKTKERAE TAELEADAR EKDKELVEAL KRLKDYESGV YGLEDAVVEI
 101 KNCKNQIKIR DREIELITKE IKKLEKISD FLDENEALRE RVGLEPTWTI
 151 DLTERNRKH LKOOQYRAEN QILKEITESL EEBRLDLKK IRQMAQERCK
 201 RSATSGLTTE DMLTENISQ GDRISERKLD LLSLKNMSEA QSKNEFLSRE
 251 LIEKREDLER SRVIAKQCN KKLKELVEENK QLEEGMKEL QAIKEMQKDP
 301 DYKGGSTSLI IPSLERLYNA IESKNAGIF DASLHLKQV DQLTGRNEEL
 351 ROELRESKKE AINYSQOLAK ANKIDHEK ETSILROSEG SNVFFKCIDL
 401 PDGIAPSSAS IINSQNEYL HLLQLEENKE KKLKNEEDSL EDYNRKFAVI
 451 RHQSLLYKE YLSEKETWKT ESKTKEER KLEDQVQDA IKVKEYNNLL
 501 NALQMSDEM KKLIAENSRK IYLYQVNEKS LIRQTTTIVE LEROLKRENE
 551 KOKNELSME AEVCEKIGCL QREKKAIFK IALQKAVDN SVSISELELA
 601 NKQYNELTAK YRDILQKDNM LVQRTSNLEH LECENISLKE QVESINKLE
 651 ITKEKLHIE QAMEQETKIG NESSMDKAKK SITNSDIVSI SKKKKKKKKK
 701 KKKKKKKKK KKKKKKKKK

11AA_SEQUENCE 1.0
 ID Q9H5V6 PRELIMINARY; PRT; 168 AA.
 AC Q9H5V6;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE CDNA: FLJ22976 FIS, CLONE KAT11222 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK026622; BAB15513.1;
 FT NOVELTER 168 AA; 19349 MW; A19DBD195F8A1A90 CRC64;
 SO SEQUENCE 168 AA; 19349 MW; A19DBD195F8A1A90 CRC64;

Q9H5V6 Length: 168 July 1, 2002 15:52 Type: P Check: 379

1 MNGNGSSGIQ QKGNVQGA ATPTASASC QYRCIECNE AKELYRQYKH
 51 GVLKTTICKS CQKPYDKYIE YDPIIILINA ILCKAQAYRH ILFNTQINIH
 101 GKLCIFCLIC EAYLRWQLOQ DSNQNTAPDD LIRYVREMEK KKKKKKKKKK
 151 KKKKKKKKK KKKKKKK

11AA_SEQUENCE 1.0
 ID Q9NT34 PRELIMINARY; PRT; 380 AA.

AC Q9NT34;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).
 GN DKFZP4341120.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Otterwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A1137556; CAB70810.1;
 KW Hypothetical protein
 FT NOVELTER 380 AA; 42689 MW; 67F50DD101346AFB CRC64;
 SO SEQUENCE 380 AA; 42689 MW; 67F50DD101346AFB CRC64;

Q9NT34 Length: 380 July 1, 2002 15:52 Type: P Check: 7330

1 MSTDSKLNK KRAYQLTK TQPEATDA FMDQFWADTA TSVQDFVALY
 51 PAAEIRAVRE ESPPSLALIC YKAVEKLYG AESGCHSEK KOIYINCRL
 101 LTRVLPIEIE PDWRFQFWS TVPGAGRGG EEDDEHARPL AESTLLAIAD
 151 LTRCPDFTVQ SHRSTVDSA EDVHSLDCE YIWEAGVQFA HSPQPNYTHD
 201 MNRMELLKLL LTCFSEAWLT PPAESGSTN PMVQFCSTN NNHALPLFTS
 251 LNTVCAVDP VGYGIPIYHL LPSDYREPLV EEAQVLIYV LDHDSASSAS
 301 PTVDGTTTGT AMDADPPGP ENLFVNTLSR IHREDFOFI LKGIARLLSN
 351 LLLQKKKKKK KKKKKKKKK

11AA_SEQUENCE 1.0
 ID Q95LV6 PRELIMINARY; PRT; 531 AA.
 AC Q95LV6;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHETICAL 61.4 KDA PROTEIN (FRAGMENT).
 OS Macaca fascicularis (Orab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheidae; Macaca.
 OX NCBI_Taxid=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
 RA Terao K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 libraries."
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB071085; BAB64479.1;
 KW Hypothetical protein.
 FT NOVELTER 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;
 SO SEQUENCE 531 AA; 61389 MW; B55996B4F5CDD60C CRC64;

Q95LV6 Length: 531 July 1, 2002 15:52 Type: P Check: 7515

1 MESESSNANM MNVQEREDK NIQKMLPSV PCTSOHLSES TYQMDPPDC
 51 KSRSEPSPE GRSSNLSHI VQKTEQETHF RESYLEPISG YMMKOSPENQ
 101 EGKCVVGLK TSPPTGKSE IGSMPHDPW DENPRKWD SISEKTANWP
 151 KNOTIVLKL DPFSLMSSEY ESRSTLEFI GKKSITSPKH VTLKTKOLPI

```

201 SOLFNIRCS TENHKKKHQ CFYKMKGRQ WYTSIGELR SATEYAKSP
251 SKSMIDKLE NTAAGTISN RTHQNVYH TTEEKEVOE NVAASSIGPL
301 DEFMPYLSDS KNQNTIRLS ERKTIINPKC LTMKEKSP1 SQIRKINHE
351 TRHKKKLES NLKTRKAMW QGENVDTFP NTISFPDTS DIKQSRFQI
401 EIDMRISGS HTQPTQISL AEGIARCSK RRTSNLYKT KLHRESEK
451 KOEHLTMDP FYAENEMTNT HLKRPDLK SEDVILGEF ISKQFYKGN
501 SKKKKKKKK KKKKKKKKK KKKKKKKKK K

11AA_SEQUENCE 1.0
ID 09LXK2: PRELIMINARY; PRT: 517 AA.
AC 09LXK2:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 18, Last annotation update)
GN HYPOTHETICAL 59.7 KDA PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA Rued S., Lemcke K., Mayer K.F.X., Quefiet F., Salanoubat M.,
RA Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL353032; CAB88307.1;
DR InterPro: IPR001810; F-box; 1.
DR Pfam: PF00646; F-box; 1.
DR SMART: SM00256; FBOX; 1.
DR PROSITE: PS00181; FBOX; 1.
DR Hypothetical protein.
KW SEQUENCE 517 AA; 59689 MW; EC6D957D01F8E670 CRC64;
SQ
09LXK2 Length: 517 July 1, 2002 15:52 Type: P Check: 4143
1 MDLFSSLPNE LYHILSFS TKRALTSVL SKRRLNPAF VYLEFDDSV
51 FLHPEERKRE KEGIIQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV
101 DRWICNVLAR GVSDDLFTD FRDLYSLPHE VGVSRLLYVL RVGSSEDLXW
151 WQFLCLPML KTLVDSCLW CIGFOIILL ACALAEELDM TTRKKSNSV
201 TVSSSIKEL TIDLHGCSV VNLKSLSPDA PSIVFYCYD SLAEDYPOVN
251 LKNLVEQIN LLTQAQIEQ VRAINNELV ADVVFGELN AKKLITGLN
301 VQQLYLSPT LEVLSRCEG MPVENNLKVL STWSDMNGW QAMPVILRNC
351 PHETLLIEG LHYATDKG DVCDISRDY KDSLSLSCP KKLQIYEFEG
401 TIRELEMIKH FLKIFPCKE MDIYAENSH TLFKDPITFE RVGKKKKKKK
451 KKKKKKKKK KKKKKIRLN FKPVAKTQF LKRLADKLCF IQQCEFLDY
501 DSSLGELALL AMDSRPS

11AA_SEQUENCE 1.0
ID 09LG29: PRELIMINARY; PRT: 260 AA.
AC 09LG29:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

```

```

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE:FLD9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA;
RA Nakamura Y.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
RL EMBL: AP002460; BAA97098.1;
DR EMBL: AP002460; BAA97098.1;
DR InterPro: IPR001385; Linker_histone.
DR PRINTS: PR00624; HISTONEH5.
SQ SEQUENCE 260 AA; 33307 MW; 43E2394CB8131143 CRC64;
09LG29 Length: 260 July 1, 2002 15:52 Type: P Check: 6094
1 MDRCIRKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
51 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
101 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
151 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
201 KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK KKKKKKKKK
251 INWAGFYIF

11AA_SEQUENCE 1.0
ID 035807: PRELIMINARY; PRT: 129 AA.
AC 035807:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MICROVASCULAR ENDOTHELIAL DIFFERENTIATION PROTEIN 2.
GN MDG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EPIDIDYMIS;
RX MEDLINE=98172708; PubMed=9511718;
RA Pirels F., Loser B., Marx M.;
RT "Differential expression of osteopontin, PC4, and CDC5, a novel mRNA
RT species, during in vitro angiogenesis.";
RL Exp. Cell Res. 239:1-10(1998).
DR EMBL: Y08769; CAAT0022.1;
DR InterPro: IPR00719; Euk_Pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 129 AA; 15080 MW; 38102272BBE2EDB4 CRC64;
035807 Length: 129 July 1, 2002 15:52 Type: P Check: 7510
1 MLKHPHVEL LETYSSDGM YWFEFMDGA DLCEFIYKRA DAGFVSEAV
51 ASHYKQILE ALRYCHDNNI IHRDVPHCV LLASKKKKK KKKKKKKKK
101 KKKKKKMG FQAFWPAIP KSSGVIIO

11AA_SEQUENCE 1.0
ID 064075: PRELIMINARY; PRT: 215 AA.
AC 064075:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

```


DE NUCLEOPORIN p52 HOMOLOG PROTEIN (FRAGMENT).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95151924; PubMed=7849178;
RA Wang Z.Q., Akmal K.M., Kim K.H.;
RT "An unusual nucleoporin-related messenger ribonucleic acid is present
in the germ cells of rat testis."
RL Biol. Reprod. 51:102-103(1994).
DR EMBL: S75997; AAB3384.1; .
KW Porin.
FT NON-TER 1 1
SQ SEQUENCE 215 AA; 24593 MW; 098251C97A8FBD88 CRC64;

O64075 Length: 215 July 1, 2002 15:52 Type: P Check: 8048 ..
1 SGGRTATSCD EDCISSLPF SLGGPVKQDC EFLEKRRKK KKKKKKKKK
51 KKKKKKTGDN AKSVRQYSL KTKLEHAE QAKVELDFIL SQKLEEDLL
101 SPLBSVKEQ SGTLYLQHAD EEREKYKLA ENIDAQLRRM AQLDKIIEH
151 LNWAGGPADT SDPLQIQICKI LNAHMDSLQW VDQSSALLQ RVEASRYCE
201 SRKEQERSL RIAFD

```

1 FINDPATTERNS on geneseqp.* allowing 0 mismatches
1 (R,K){20} July 1, 2002 06:23 ..

1 AAP20159 ck: 5750 len: 20 1 Aap20159 Sequence of lysine polymer. 8/1992
1: (R,K){20}
(K){20}
KKKKKKKKKKKKKKKKKKKK

1 AAP61030 ck: 9157 len: 898 1 Aap61030 Entire coded sequence from plasmid
(R,K){20}
(K){20}
873: KNITW KKKKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
874: NITWK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
875: ITWKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
876: TWKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
877: WKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
878: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
879: KKKKK KKKKKKKKKKKKKKKKKKKKK

1 AAP61056 ck: 2017 len: 899 1 Aap61056 Translation of plasmid PAU157 endc
(R,K){20}
873: KNITW KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
874: NITWK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
875: ITWKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
876: TWKKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
877: WKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
878: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
879: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
880: KKKKK KKKKKKKKKKKKKKKKKKKKK

1 AAP61082 ck: 7915 len: 898 1 Aap61082 Complete translation of plasmid PA
(R,K){20}
873: KNITW KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
874: NITWK KKKKKKKKKKKKKKKKKKKKK KKKKK

```

```

1 875: ITWKK KKKKKKKKKKKKKKKKKKKKK KKKK
(K){20}
876: TWKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}
877: WKKKK KKKKKKKKKKKKKKKKKKKKK KK
(K){20}
878: KKKKK KKKKKKKKKKKKKKKKKKKKK K
(K){20}
879: KKKKK KKKKKKKKKKKKKKKKKKKKK

1 AAR29580 ck: 4341 len: 657 1 Aar29580 FWR-1 gene product. 4/1993
(R,K){20}
19: RRRRP RRRRRRRRRRRRRRRRRRRR RRRRL
(R){20}
20: RRRRP RRRRRRRRRRRRRRRRRRRR RRRLG
(R){20}
21: RRP RP RRRRRRRRRRRRRRRRRRR RRLGL
(R){20}
22: RPPRR RRRRRRRRRRRRRRRRRRR RRLGLE
(R){20}
23: PRRRR RRRRRRRRRRRRRRRRRRR RLGLR
(R){20}

1 AAW03642 ck: 9623 len: 116 1 Aaw03642 Human cannabinoid GPR N-termina
(R,K){20}
34: QYEDI KKKKKKKKKKKKKKKKKKKKK KSPFQ
(K){20}
35: YEDIK KKKKKKKKKKKKKKKKKKKKK SPFOE
(K){20}

1 AAW38839 ck: 801 len: 28 1 Aaw38839 Delivery peptide used in peptid
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}

1 AAW38840 ck: 2989 len: 29 1 Aaw38840 Delivery peptide used in peptid

```

```
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK XK
  (K){20}

AAW38841 ck: 5252 len: 30 1 Aaw38841 Delivery peptide used in peptide m
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKK
  (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK XK
  (K){20}

AAW38842 ck: 7590 len: 31 1 Aaw38842 Delivery peptide used in peptide m
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
```

```
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKK
  (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK XK
  (K){20}

AAW38877 ck: 1129 len: 23 1 Aaw38877 Delivery peptide used in pepti.
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK VTK
  (K){20}

AAW38843 ck: 3 len: 32 1 Aaw38843 Delivery peptide used in pepti.
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
  (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKK
  (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK XK
  (K){20}

AAW38878 ck: 2949 len: 24 1 Aaw38878 Delivery peptide used in pepti.
1
1: (R,K){20}
  KKKKKKKKKKKKKKKKKKKKK KVTK
  (K){20}
2: K KKKKKKKKKKKKKKKKKKK VTK
  (K){20}

AAW38844 ck: 2491 len: 33 1 Aaw38844 Delivery peptide used in pepti.
```

1

(R,K){20}
 (K){20}
 1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 2: K KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKK XK

AAW38879 ck: 4844 len: 25 1 Aaw38879 Delivery peptide used in peptide m

1

(R,K){20}
 (K){20}
 1: KKKKKKKKKKKKKKKKKKKKK KKVTK

(K){20}
 2: K KKKKKKKKKKKKKKKKKKK KVTK

(K){20}
 3: KK KKKKKKKKKKKKKKKKKKK VTK

AAW38845 ck: 5054 len: 34 1 Aaw38845 Delivery peptide used in peptide m

1

(R,K){20}
 (K){20}
 1: KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 2: K KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

(K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKK
 (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKK XK

AAW38880 ck: 6814 len: 26 1 Aaw38880 Delivery peptide used in peptid

1

(R,K){20}
 (K){20}
 1: KKKKKKKKKKKKKKKKKKKKK KKVTK
 (K){20}
 2: K KKKKKKKKKKKKKKKKKKK KKVTK
 (K){20}
 3: KK KKKKKKKKKKKKKKKKKKK KVTK
 (K){20}
 4: KKK KKKKKKKKKKKKKKKKKKK VTK

AAW38846 ck: 7692 len: 35 1 Aaw38846 Delivery peptide used in peptid

(R,K){20}
 (K){20}
 1: KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 2: K KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 3: KK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 4: KKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 5: KKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

```
(K){20}
12: KKKK KKKKKKKKKKKKKKK KKK
(K){20}
13: KKKK KKKKKKKKKKKKKKK KKK
(K){20}
14: KKKK KKKKKKKKKKKKKKK KKK

AAW38633 ck: 9248 len: 22 1 Aaw38633 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK

AAW38634 ck: 986 len: 23 1 Aaw38634 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK

AAW38635 ck: 2799 len: 24 1 Aaw38635 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
(K){20}

AAW38636 ck: 4687 len: 25 1 Aaw38636 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKK

AAW38637 ck: 6650 len: 26 1 Aaw38637 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
4: KKK KKKKKKKKKKKKKKKKK KKK
(K){20}

(K){20}
5: KKK KKKKKKKKKKKKKKKKK KKK

AAW38638 ck: 8688 len: 27 1 Aaw38638 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
4: KKK KKKKKKKKKKKKKKKKK KKK
5: KKK KKKKKKKKKKKKKKKKK KKK
6: KKKK KKKKKKKKKKKKKKKKK KKK
(K){20}

AAW38796 ck: 9227 len: 22 1 Aaw38796 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK

AAW38797 ck: 964 len: 23 1 Aaw38797 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
(K){20}

AAW38798 ck: 2776 len: 24 1 Aaw38798 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
(K){20}

AAW38799 ck: 4663 len: 25 1 Aaw38799 Delivery peptide used in peptide m
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKK KKK
2: K KKKKKKKKKKKKKKKKK KKK
3: K KKKKKKKKKKKKKKKKK KKK
4: KKK KKKKKKKKKKKKKKKKK KKK
(K){20}

AAW38800 ck: 6625 len: 26 1 Aaw38800 Delivery peptide used in peptide m
(R,K){20}
```

1

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

AAW38801 ck: 8662 len: 27 1 Aaw38801 Delivery peptide used in peptide m

1

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

AAW38802 ck: 774 len: 28 1 Aaw38802 Delivery peptide used in peptide m

1

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

AAW38803 ck: 2961 len: 29 1 Aaw38803 Delivery peptide used in peptide m

1

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

1

```

3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

AAW38804 ck: 5223 len: 30 1 Aaw38804 Delivery peptide used in peptide m

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK*
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

AAW38805 ck: 7560 len: 31 1 Aaw38805 Delivery peptide used in peptide m

1

```

1: (R,K){20}
   KKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

8: KKKKK (K){20} KKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKK WK

AAM38806 ck: 9972 len: 32 ! Aaw38806 Delivery peptide used in peptide m

1: (R,K){20}
(K){20} KKKKKKKKKKKKKKKKKKK KKKKK

2: (K){20} K KKKKKKKKKKKKKKKKKKK KKKKK

3: K KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKK WK
(K){20}

AAM38807 ck: 2459 len: 33 ! Aaw38807 Delivery peptide used in peptide m

1: (R,K){20}
(K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

2: (K){20} K KKKKKKKKKKKKKKKKKKKKK KKKKK

3: (K){20} KK KKKKKKKKKKKKKKKKKKKKK KKKKK

4: (K){20} KKK KKKKKKKKKKKKKKKKKKKKK KKKKK

5: (K){20} KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

6: (K){20} KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

7: (K){20} KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK WK
(K){20}

AAM38808 ck: 5021 len: 34 ! Aaw38808 Delivery peptide used in peptide m

1: (R,K){20}
(K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

2: (K){20} K KKKKKKKKKKKKKKKKKKKKK KKKKK

3: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK WK
(K){20}

AAM38881 ck: 8859 len: 27 ! Aaw38881 Delivery peptide used in peptide m

1: (R,K){20}
(K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

2: (K){20} K KKKKKKKKKKKKKKKKKKKKK KKKKK

3: (K){20} KK KKKKKKKKKKKKKKKKKKKKK KKKKK

4: (K){20} KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

5: KKKK (K){20}
KKKKKKKKKKKKKKKKKK VTK

AAW38847 ck: 405 len: 36 ! Aaw38847 Delivery peptide used in peptide m

1: (R,K){20}
(K){20}
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKK

15: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK XK

AAW38862 ck: 979 len: 28 ! Aaw38862 Delivery peptide used in peptide m

1: (R,K){20}
(K){20}
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}
KKKKKKKKKKKKKKKKKK KKKVT

4: KKK (K){20}
KKKKKKKKKKKKKKKKKK KKVTK

5: KKKK (K){20}
KKKKKKKKKKKKKKKKKK KVTK

6: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK VTK

AAW38848 ck: 3193 len: 37 ! Aaw38848 Delivery peptide used in peptide m

1: (R,K){20}
(K){20}
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

7: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

8: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

9: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

10: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

11: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

12: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

13: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

14: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

15: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKK

16: KKKKKK (K){20}
KKKKKKKKKKKKKKKKKK XK

AAW38863 ck: 3174 len: 29 ! Aaw38863 Delivery peptide used in peptide m

1: (R,K){20}
(K){20}
KKKKKKKKKKKKKKKKKK KKKKK

2: K (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

3: KK (K){20}
KKKKKKKKKKKKKKKKKK KKKVT

4: KKK (K){20}
KKKKKKKKKKKKKKKKKK KKVTK

5: KKKK (K){20}
KKKKKKKKKKKKKKKKKK KVTK

6: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK VTK

6: KKKKK KKKKKKKKKKKKKKKKKKKKK KVTk

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK

AAW38849 ck: 6056 len: 38 ! Aaw38849 Delivery peptide used in peptide m

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38884 ck: 5444 len: 30 ! Aaw38884 Delivery peptide used in peptide m

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

AAW38850 ck: 8994 len: 39 ! Aaw38850 Delivery peptide used in pepti

1

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}


```
6: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKK KKK
   (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKK KK
   (K){20}

AAW38887 ck: 2704 len: 33 ! Aaw38887 Delivery peptide used in peptide m
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
6: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKKKK KKKV
   (K){20}
   (K){20}
```

```
9: KKKK KKKKKKKKKKKKKKKKKKK KKKV
   (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKV
   (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK VTK
   (K){20}

AAW38853 ck: 8258 len: 42 ! Aaw38853 Delivery peptide used in peptide
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
6: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
9: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
10: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
11: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
12: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
13: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
14: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
15: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
16: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
17: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
18: KKKK KKKKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
19: KKKK KKKKKKKKKKKKKKKKKKKKK KKK
   (K){20}
20: KKKK KKKKKKKKKKKKKKKKKKKKK KKK
   (K){20}
```

21: KKKK KKKKKKKKKKKKKKKKKKK X

AAW38888 ck: 5274 len: 34 ! Aaw38888 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

AAW38854 ck: 1496 len: 43 ! Aaw38854 Delivery peptide used in peptide m

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

18: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

19: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

20: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

21: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

22: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

AAW38889 ck: 7919 len: 35 ! Aaw38889 Delivery peptide used in peptid

(R,K){20}

1: KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

2: K KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

3: KK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

4: KKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

5: KKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

7: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

8: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

9: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: KKKKKK KKKKKKKKKKKKKKKKKKKKKKK

11: KKKK (K){20} KKVTK
12: KKKK (K){20} KKVTK
13: KKKK (K){20} VTK

AAW38890 ck: 639 len: 36 ! Aaw38890 Delivery peptide used in peptide m

1

(R,K){20}
1: KKKK (K){20} KKKK

2: K KKKK (K){20} KKKK

3: KK KKKK (K){20} KKKK

4: KK KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKVTK

12: KKKK (K){20} KKVTK

13: KKKK (K){20} KVT

14: KKKK (K){20} VTK

AAW38891 ck: 3434 len: 37 ! Aaw38891 Delivery peptide used in peptide m

1

(R,K){20}
1: KKKK (K){20} KKKK

2: K KKKK (K){20} KKKK

3: KK KKKK (K){20} KKKK

4: KK KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

-1

AAW38892 ck: 6304 len: 38 ! Aaw38892 Delivery peptide used in pepti

1

(R,K){20}
1: KKKK (K){20} KKKK

2: K KKKK (K){20} KKKK

3: KK KKKK (K){20} KKKK

4: KK KKKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKKK

12: KKKK (K){20} KKKK

13: KKKK (K){20} KKVTK

14: KKKK (K){20} KKVTK
15: KKKK (K){20} KKVTK
16: KKKK (K){20} VTK

AAW38893 ck: 9249 len: 39 1 Aaw38893 Delivery peptide used in peptide m

1

(R,K){20}
(K){20}
1: KKKK

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKKK

12: KKKK (K){20} KKKK

13: KKKK (K){20} KKKK

14: KKKK (K){20} KKVTK

15: KKKK (K){20} KKVTK

16: KKKK (K){20} KVTK

17: KKKK (K){20} VTK

AAW38894 ck: 2269 len: 40 1 Aaw38894 Delivery peptide used in peptide m

1

(R,K){20}
(K){20}
1: KKKK

(K){20}

2: K KKKK

3: KK (K){20} KKKK

4: KKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK

9: KKKK (K){20} KKKK

10: KKKK (K){20} KKKK

11: KKKK (K){20} KKKK

12: KKKK (K){20} KKKK

13: KKKK (K){20} KKKK

14: KKKK (K){20} KKVTK

15: KKKK (K){20} KKVTK

16: KKKK (K){20} KKVTK

17: KKKK (K){20} KVTK

18: KKKK (K){20} VTK

AAW38895 ck: 5364 len: 41 1 Aaw38895 Delivery peptide used in peptid

1

(R,K){20}
(K){20}
1: KKKK

2: K (K){20} KKKK

3: KK (K){20} KKKK

4: KKK (K){20} KKKK

5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK

(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAW38896 ck: 8534 len: 42 1 Aaw38896 Delivery peptide used in peptide m
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAW38897 ck: 1779 len: 43 1 Aaw38897 Delivery peptide used in pepti.
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

```

14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK
    (K){20}

AAW38898 ck: 5099 len: 44 ! Aaw38898 Delivery peptide used in peptide m
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}

```

```

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAW38809 ck: 7658 len: 35 ! Aaw38809 Delivery peptide used in peptid
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKKKK KKKKK
2: K KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
  (K){20}

AAW38810 ck: 370 len: 36 ! Aaw38810 Delivery peptide used in peptid
1: (R,K){20}
  (K){20}
  KKKKKKKKKKKKKKKKKKKKKKK KKKKK

```


1
2: K (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
3: KK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
4: KKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
5: KKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAW38811 ck: 3157 len: 37 1 Aaw38811 Delivery peptide used in peptide m

1
10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAW38812 ck: 6019 len: 38 1 Aaw38812 Delivery peptide used in pepti

17: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK WK

AAW38813 ck: 8956 len: 39 ! Aaw38813 Delivery peptide used in peptide m

1: (R,K)[20]
(K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

AAW38814 ck: 1968 len: 40 ! Aaw38814 Delivery peptide used in peptide m

1: (R,K)[20]
(K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

9: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

10: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

11: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

12: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

13: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

14: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

15: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

16: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

17: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

18: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

19: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

AAW38815 ck: 5055 len: 41 ! Aaw38815 Delivery peptide used in peptide m

1: (R,K)[20]
(K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

2: K (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

3: KK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

4: KKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

5: KKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

6: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

7: KKKKK (K)[20]
KKKKKKKKKKKKKKKKKKKK KKKKK

8: KKKK (K){20} KKKK
9: KKKK (K){20} KKKK
10: KKKK (K){20} KKKK
11: KKKK (K){20} KKKK
12: KKKK (K){20} KKKK
13: KKKK (K){20} KKKK
14: KKKK (K){20} KKKK
15: KKKK (K){20} KKKK
16: KKKK (K){20} KKKK
17: KKKK (K){20} KKKK
18: KKKK (K){20} KKKK
19: KKKK (K){20} KKKK
20: KKKK (K){20} KKKK WK
AAW38816 ck: 8217 len: 42 1 Aaw38816 Delivery peptide used in peptide p
(R,K){20}
1: KKKK (K){20} KKKK
2: K (K){20} KKKK
3: K (K){20} KKKK
4: K (K){20} KKKK
5: K (K){20} KKKK
6: K (K){20} KKKK
7: K (K){20} KKKK
8: K (K){20} KKKK
9: K (K){20} KKKK
10: K (K){20} KKKK

11: KKKK (K){20} KKKK
12: KKKK (K){20} KKKK
13: KKKK (K){20} KKKK
14: KKKK (K){20} KKKK
15: KKKK (K){20} KKKK
16: KKKK (K){20} KKKK
17: KKKK (K){20} KKKK
18: KKKK (K){20} KKKK
19: KKKK (K){20} KKKK
20: KKKK (K){20} KKKK WK
21: KKKK (K){20} KKKK WK
AAW38817 ck: 1454 len: 43 1 Aaw38817 Delivery peptide used in peptic
(R,K){20}
1: KKKK (K){20} KKKK
2: K (K){20} KKKK
3: K (K){20} KKKK
4: K (K){20} KKKK
5: K (K){20} KKKK
6: K (K){20} KKKK
7: K (K){20} KKKK
8: K (K){20} KKKK
9: K (K){20} KKKK
10: K (K){20} KKKK
11: K (K){20} KKKK
12: K (K){20} KKKK

(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
AAW24865 ck: 2211 len: 40 ! Aaw24865 Bifunctional peptide I for binding
(R,K){20}
21: YEDES KKKKKKKKKKKKKKKKKKK KKKKK
AAW24450 ck: 8137 len: 45 ! Aaw24450 Nucleic acid (NA) binding peptide
(R,K){20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: YKAK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: YKAKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
AAW21590 ck: 4875 len: 30 ! Aaw21590 Antibiotic potentiating peptide
(R,K){20}
1: KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAW21591 ck: 5075 len: 434 ! Aaw21591 Antibiotic potentiating peptide
(R,K){20}

50: KKKK (K){20} KKKK
 51: KKKK (K){20} KKKK
 52: KKKK (K){20} KKKK
 53: KKKK (K){20} KKKK
 54: KKKK (K){20} KKKK
 55: KKKK (K){20} KKKK
 56: KKKK (K){20} KKKK
 57: KKKK (K){20} KKKK
 58: KKKK (K){20} KKKK
 59: KKKK (K){20} KKKK
 60: KKKK (K){20} KKKK
 61: KKKK (K){20} KKKK
 62: KKKK (K){20} KKKK
 63: KKKK (K){20} KKKK
 64: KKKK (K){20} KKKK
 65: KKKK (K){20} KKKK
 66: KKKK (K){20} KKKK
 67: KKKK (K){20} KKKK
 68: KKKK (K){20} KKKK
 69: KKKK (K){20} KKKK
 70: KKKK (K){20} KKKK
 71: KKKK (K){20} KKKK
 72: KKKK (K){20} KKKK
 73: KKKK (K){20} KKKK

74: KKKK (K){20} KKKK
 75: KKKK (K){20} KKKK
 76: KKKK (K){20} KKKK
 77: KKKK (K){20} KKKK
 78: KKKK (K){20} KKKK
 79: KKKK (K){20} KKKK
 80: KKKK (K){20} KKKK
 81: KKKK (K){20} KKKK
 82: KKKK (K){20} KKKK
 83: KKKK (K){20} KKKK
 84: KKKK (K){20} KKKK
 85: KKKK (K){20} KKKK
 86: KKKK (K){20} KKKK
 87: KKKK (K){20} KKKK
 88: KKKK (K){20} KKKK
 89: KKKK (K){20} KKKK
 90: KKKK (K){20} KKKK
 91: KKKK (K){20} KKKK
 92: KKKK (K){20} KKKK
 93: KKKK (K){20} KKKK
 94: KKKK (K){20} KKKK
 95: KKKK (K){20} KKKK
 96: KKKK (K){20} KKKK
 97: KKKK (K){20} KKKK

98: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
99: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
100: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
101: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
102: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
103: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
104: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
105: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
106: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
107: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
108: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
109: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
110: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
111: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
112: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
113: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
114: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
115: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
116: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
117: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
118: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
119: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
120: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
121: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
122: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}

123: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
124: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
125: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
126: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
127: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
128: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
129: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
130: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
131: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
132: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
133: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
134: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
135: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
136: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
137: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
138: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
139: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
140: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
141: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
142: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
143: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
144: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
145: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
146: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}

147: KKKK (K){20} KKKK
 148: KKKK (K){20} KKKK
 149: KKKK (K){20} KKKK
 150: KKKK (K){20} KKKK
 151: KKKK (K){20} KKKK
 152: KKKK (K){20} KKKK
 153: KKKK (K){20} KKKK
 154: KKKK (K){20} KKKK
 155: KKKK (K){20} KKKK
 156: KKKK (K){20} KKKK
 157: KKKK (K){20} KKKK
 158: KKKK (K){20} KKKK
 159: KKKK (K){20} KKKK
 160: KKKK (K){20} KKKK
 161: KKKK (K){20} KKKK
 162: KKKK (K){20} KKKK
 163: KKKK (K){20} KKKK
 164: KKKK (K){20} KKKK
 165: KKKK (K){20} KKKK
 166: KKKK (K){20} KKKK
 167: KKKK (K){20} KKKK
 168: KKKK (K){20} KKKK
 169: KKKK (K){20} KKKK
 170: KKKK (K){20} KKKK

171: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 172: KKKK (K){20} KKKK
 173: KKKK (K){20} KKKK
 174: KKKK (K){20} KKKK
 175: KKKK (K){20} KKKK
 176: KKKK (K){20} KKKK
 177: KKKK (K){20} KKKK
 178: KKKK (K){20} KKKK
 179: KKKK (K){20} KKKK
 180: KKKK (K){20} KKKK
 181: KKKK (K){20} KKKK
 182: KKKK (K){20} KKKK
 183: KKKK (K){20} KKKK
 184: KKKK (K){20} KKKK
 185: KKKK (K){20} KKKK
 186: KKKK (K){20} KKKK
 187: KKKK (K){20} KKKK
 188: KKKK (K){20} KKKK
 189: KKKK (K){20} KKKK
 190: KKKK (K){20} KKKK
 191: KKKK (K){20} KKKK
 192: KKKK (K){20} KKKK
 193: KKKK (K){20} KKKK
 194: KKKK (K){20} KKKK
 195: KKKK (K){20} KKKK

196: KKKK (K){20} KKKK
197: KKKK (K){20} KKKK
198: KKKK (K){20} KKKK
199: KKKK (K){20} KKKK
200: KKKK (K){20} KKKK
201: KKKK (K){20} KKKK
202: KKKK (K){20} KKKK
203: KKKK (K){20} KKKK
204: KKKK (K){20} KKKK
205: KKKK (K){20} KKKK
206: KKKK (K){20} KKKK
207: KKKK (K){20} KKKK
208: KKKK (K){20} KKKK
209: KKKK (K){20} KKKK
210: KKKK (K){20} KKKK
211: KKKK (K){20} KKKK
212: KKKK (K){20} KKKK
213: KKKK (K){20} KKKK
214: KKKK (K){20} KKKK
215: KKKK (K){20} KKKK
216: KKKK (K){20} KKKK
217: KKKK (K){20} KKKK
218: KKKK (K){20} KKKK
219: KKKK (K){20} KKKK

9

220: KKKK (K){20} KKKK
221: KKKK (K){20} KKKK
222: KKKK (K){20} KKKK
223: KKKK (K){20} KKKK
224: KKKK (K){20} KKKK
225: KKKK (K){20} KKKK
226: KKKK (K){20} KKKK
227: KKKK (K){20} KKKK
228: KKKK (K){20} KKKK
229: KKKK (K){20} KKKK
230: KKKK (K){20} KKKK
231: KKKK (K){20} KKKK
232: KKKK (K){20} KKKK
233: KKKK (K){20} KKKK
234: KKKK (K){20} KKKK
235: KKKK (K){20} KKKK
236: KKKK (K){20} KKKK
237: KKKK (K){20} KKKK
238: KKKK (K){20} KKKK
239: KKKK (K){20} KKKK
240: KKKK (K){20} KKKK
241: KKKK (K){20} KKKK
242: KKKK (K){20} KKKK
243: KKKK (K){20} KKKK

244: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 245: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 246: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 247: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 248: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 249: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 250: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 251: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 252: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 253: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 254: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 255: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 256: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 257: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 258: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 259: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 260: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 261: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 262: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 263: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 264: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 265: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 266: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 267: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 268: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}

269: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 270: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 271: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 272: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 273: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 274: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 275: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 276: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 277: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 278: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 279: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 280: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 281: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 282: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 283: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 284: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 285: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 286: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 287: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 288: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 289: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 290: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 291: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 292: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}

293: KKKK (K){20} KKKK
294: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
295: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
296: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
297: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
298: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
299: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
300: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
301: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
302: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
303: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
304: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
305: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
306: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
307: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
308: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
309: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
310: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
311: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
312: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
313: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
314: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
315: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
316: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK

317: KKKK KKKKKKKKKKKKKKKKK KKKK
318: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
319: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
320: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
321: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
322: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
323: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
324: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
325: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
326: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
327: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
328: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
329: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
330: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
331: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
332: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
333: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
334: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
335: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
336: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
337: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
338: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
339: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
340: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK
341: KKKK (K){20} KKKK KKKKKKKKKKKKKKKKK

342: KKKK (K){20} KKKK
 343: KKKK (K){20} KKKK
 344: KKKK (K){20} KKKK
 345: KKKK (K){20} KKKK
 346: KKKK (K){20} KKKK
 347: KKKK (K){20} KKKK
 348: KKKK (K){20} KKKK
 349: KKKK (K){20} KKKK
 350: KKKK (K){20} KKKK
 351: KKKK (K){20} KKKK
 352: KKKK (K){20} KKKK
 353: KKKK (K){20} KKKK
 354: KKKK (K){20} KKKK
 355: KKKK (K){20} KKKK
 356: KKKK (K){20} KKKK
 357: KKKK (K){20} KKKK
 358: KKKK (K){20} KKKK
 359: KKKK (K){20} KKKK
 360: KKKK (K){20} KKKK
 361: KKKK (K){20} KKKK
 362: KKKK (K){20} KKKK
 363: KKKK (K){20} KKKK
 364: KKKK (K){20} KKKK
 365: KKKK (K){20} KKKK

366: KKKK (K){20} KKKK
 367: KKKK (K){20} KKKK
 368: KKKK (K){20} KKKK
 369: KKKK (K){20} KKKK
 370: KKKK (K){20} KKKK
 371: KKKK (K){20} KKKK
 372: KKKK (K){20} KKKK
 373: KKKK (K){20} KKKK
 374: KKKK (K){20} KKKK
 375: KKKK (K){20} KKKK
 376: KKKK (K){20} KKKK
 377: KKKK (K){20} KKKK
 378: KKKK (K){20} KKKK
 379: KKKK (K){20} KKKK
 380: KKKK (K){20} KKKK
 381: KKKK (K){20} KKKK
 382: KKKK (K){20} KKKK
 383: KKKK (K){20} KKKK
 384: KKKK (K){20} KKKK
 385: KKKK (K){20} KKKK
 386: KKKK (K){20} KKKK
 387: KKKK (K){20} KKKK
 388: KKKK (K){20} KKKK
 389: KKKK (K){20} KKKK

390: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
391: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
392: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
393: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
394: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
395: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
396: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
397: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
398: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
399: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
400: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
401: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
402: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
403: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
404: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
405: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
406: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
407: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
408: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
409: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
410: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
411: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
412: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
413: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
414: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

415: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
AAW5939 ck: 1569 len: 40 1 Aaw5939 Polyllysine peptide NBC32. 11/15
1
(R,K){20}
2: T KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
3: TK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
4: TRK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
5: TKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
6: TKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
7: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
8: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
9: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
10: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
11: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
12: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
13: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
14: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
16: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
17: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
18: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
AAW4808 ck: 9658 len: 56 1 Aaw4808 Homo sapiens clone CG109_1 prot
(R,K){20}
34: EEPRE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
35: EPREK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
36: FREKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

37: REKKK KKKKKKKKKKKKKKKKKK
1 AAM45801 ck: 9500 len: 39 | Aaa45801 One chain of a bombesin dimer. 6/1
(R,K){20}
(K){20}
2: C KKKKKKKKKKKKKKKKKKK XGCGQ
1 AAM45800 ck: 9056 len: 39 | Aaa45800 One chain of an alpha-melanocyte s
(R,K){20}
(K){20}
18: VGGGX KKKKKKKKKKKKKKKKKK CX
1 AAM45802 ck: 7664 len: 35 | Aaa45802 One chain of an alpha-MSH receptor
(R,K){20}
(K){20}
14: VGGGX KKKKKKKKKKKKKKKKKK CX
1 AAY43246 ck: 9752 len: 32 | Aay43246 Cell-surface molecule binding pep
(R,K){20}
(K){20}
12: SGSGS KKKKKKKKKKKKKKKKKK K
(K){20}
13: GSGSK KKKKKKKKKKKKKKKKKK
1 AAY07213 ck: 2211 len: 40 | Aay07213 Peptide transfection vector #1. 7/
(R,K){20}
(K){20}
21: YEDES KKKKKKKKKKKKKKKKKK
1 AAY12950 ck: 5821 len: 62 | Aay12950 Amino acid sequence of a human seq
(R,K){20}
(K){20}
40: QFOAS KKKKKKKKKKKKKKKKKK KKK
(K){20}
41: FOASK KKKKKKKKKKKKKKKKKK KK
(K){20}
42: QASKK KKKKKKKKKKKKKKKKKK K
(K){20}
43: ASKKK KKKKKKKKKKKKKKKKKK
1 AAB59105 ck: 8456 len: 27 | Aab59105 Breast and ovarian cancer associat
(R,K){20}
(K){20}
6: NSAXX KKKKKKKKKKKKKKKKKK KK
(K){20}
7: SAXXX KKKKKKKKKKKKKKKKKK K
(K){20}
8: AXKKK KKKKKKKKKKKKKKKKKK
1 AAB53249 ck: 4945 len: 59 | Aab53249 Human colon cancer antigen protein
(R,K){20}

27: ALSLK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
28: LLSLK KKKKKKKKKKKKKKKKKK KKKKN (K){20}
29: LSLKK KKKKKKKKKKKKKKKKKK KKKXX (K){20}
30: SLKKK KKKKKKKKKKKKKKKKKK KKNXX (K){20}
31: LKKKK KKKKKKKKKKKKKKKKKK KXXGG (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKK NXGGG (K){20}
1 AAB53659 ck: 3850 len: 184 | Aab53659 Human colon cancer antigen prot
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKK KKKKK (K){20}
3: KK KKKKKKKKKKKKKKKKKK KKKXX (K){20}
4: KKK KKKKKKKKKKKKKKKKKK KXXGG (K){20}
5: KKKK KKKKKKKKKKKKKKKKKK KXXGR (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKK XGGRF (K){20}
1 AAB53800 ck: 296 len: 69 | Aab53800 Human colon cancer antigen prot
(R,K){20}
(K){20}
20: NYDCE KKKKKKKKKKKKKKKKKK KKKKK (K){20}
21: VDCEK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
22: DCEKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
23: CEKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
24: EKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKK KKKKK (K){20}

```

30: KKKK (K){20}
31: KKKK (K){20}
32: KKKK (K){20}
33: KKKK (K){20}
34: KKKK (K){20}
35: KKKK (K){20}
36: KKKK (K){20}
37: KKKK (K){20}
38: KKKK (K){20}
39: KKKK (K){20}
40: KKKK (K){20}

AAB53806 ck: 8373 len: 64 ! Aab53806 Human colon cancer antigen protein
(R,K){20}
25: IDCS KKKK (K){20}
26: DCDSK KKKK (K){20}
27: CDSKK KKKK (K){20}
28: DSXKK KKKK (K){20}
29: SKKKK KKKK (K){20}
30: KKKKK KKKK (K){20}

AAB53977 ck: 7756 len: 75 ! Aab53977 Human colon cancer antigen protein
(R,K){20}
24: QFCLL KKKK (K){20}
25: FCLLK KKKK (K){20}
26: CLKKK KKKK (K){20}
27: LKKKK KKKK (K){20}

```

```

28: LKKKK KKKK
29: KKKKK KKKK
30: KKKKK KKKK
31: KKKKK KKKK
32: KKKKK KKKK
33: KKKKK KKKK
34: KKKKK KKKK
35: KKKKK KKKK
36: KKKKK KKKK
37: KKKKK KKKK
38: KKKKK KKKK
39: KKKKK KKKK
40: KKKKK KKKK
41: KKKKK KKKK
42: KKKKK KKKK
43: KKKKK KKKK
44: KKKKK KKKK
45: KKKKK KKKK
46: KKKKK KKKK
47: KKKKK KKKK
48: KKKKK KKKK
49: KKKKK KKKK

AAB53980 ck: 881 len: 45 ! Aab53980 Human colon cancer antigen pro
(R,K){20}
8: ICSFQ KKKK (K){20}

```

9: CSFQK KKKKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
10: SPQK KKKKKKKKKKKKKKKKKKKKK KXXGG
(K){20}
11: FQKK KKKKKKKKKKKKKKKKKKKKK KXGGR
(K){20}
12: QKKK KKKKKKKKKKKKKKKKKKKKK XGGRF
(K){20}
AAB54314 ck: 6235 len: 55 1 Aab54314 Human pancreatic cancer antigen pr
(R,K){20}
24: FFXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: FXXXX KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: XXXKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: XKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGRF
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK GGRF
(K){20}
AAB56121 ck: 5941 len: 125 1 Aab56121 Human secreted protein sequence en
(R,K){20}
87: ILKSQ KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
88: LKSQK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
89: KSQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
90: SQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
91: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGR
(K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGP
(K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK GGP
(K){20}
AAB27956 ck: 6732 len: 139 1 Aab27956 Human secreted protein SEQ ID N
(R,K){20}
91: HPRPG KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
92: PRPGK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
93: RPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
94: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: GKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

106: KKKK (K){20} KKKK
107: KKKK (K){20} KKKK
108: KKKK (K){20} KKKK
109: KKKK (K){20} KKKK
110: KKKK (K){20} KKKK
111: KKKK (K){20} KKKK
112: KKKK (K){20} KKKK
113: KKKK (K){20} KKKK
114: KKKK (K){20} KKKK
115: KKKK (K){20} KKKK
116: KKKK (K){20} KKKK
117: KKKK (K){20} KKKK
118: KKKK (K){20} KKKK
119: KKKK (K){20} KKKK
120: KKKK (K){20} KKKK
AAB39140 ck: 9531 len: 66 1 Aab39140 Human secreted protein #48. 2/2001
(R,K){20}
41: MWTVX KKKK (K){20} KKKK
42: WTVXK KKKK (K){20} KKKK
43: TVXKK KKKK (K){20} KKKK
44: VXKKK KKKK (K){20} KKKK
45: XKKKK KKKK (K){20} KKKK
46: KKKKK KKKK (K){20} KKKK
47: KKKKK KKKK (K){20} KKKK
AAB41457 ck: 4536 len: 168 1 Aab41457 Human ORF1221 polypeptide seq

144: KQKRP KKKK (R,K){20} KKKK
145: QQKPK KKKK (K){20} KKKK
146: QKPKK KKKK (K){20} KKKK
147: KPKKK KKKK (K){20} KKKK
148: PKKKK KKKK (K){20} KKKK
149: KKKKK KKKK (K){20} KKKK
AAB42786 ck: 4959 len: 102 1 Aab42786 Human ORF2550 polypeptide
(R,K){20}
81: SDVLQ KKKK (K){20} KKKK
82: DVLQK KKKK (K){20} KKKK
83: VLQKK KKKK (K){20} KKKK
AAB42889 ck: 6432 len: 62 1 Aab42889 Human ORF2653 polypeptide
(R,K){20}
40: SHRCL KKKK (K){20} KKKK
41: HRCLK KKKK (K){20} KKKK
42: RCLCK KKKK (K){20} KKKK
43: CLKKK KKKK (K){20} KKKK
AAB43641 ck: 5216 len: 133 1 Aab43641 Human cancer associated protei
(R,K){20}
114: GHEQS KKKK (K){20} KKKK
AAB43835 ck: 4025 len: 223 1 Aab43835 Human cancer associated protei
(R,K){20}
196: NILFW KKKK (K){20} KKKK
AAB44188 ck: 4991 len: 43 1 Aab44188 Human cancer associated protei
(R,K){20}
2: E KKKK (K){20} KKKK
3: EK KKKK (K){20} KKKK

```

4:  EKK KKKKKKKKKKKKKKKKKKKKK KKKGR
   (K){20}
5:  EKKK KKKKKKKKKKKKKKKKKKKKK KKGSP
   (K){20}
6:  EKKKK KKKKKKKKKKKKKKKKKKKKK KGPEF
   (K){20}
7:  KKKKK KKKKKKKKKKKKKKKKKKKKK GRPEX
   (K){20}

AAB44380  ck: 8490  len: 42  ! Aab44380 Human secreted protein encoded by
   (R,K){20}
22:  IWKKI KKKKKKKKKKKKKKKKKKKKK K
   (K){20}
23:  WKIKI KKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAB43331  ck: 8131  len: 66  ! Aab43331 Human secreted protein sequence en
   (R,K){20}
40:  IHMKP KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
41:  HHMKP KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
42:  HKPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
43:  KPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}
44:  PKKKK KKKKKKKKKKKKKKKKKKKKK KKK
   (K){20}
45:  KKKKK KKKKKKKKKKKKKKKKKKKKK KK
   (K){20}
46:  KKKKK KKKKKKKKKKKKKKKKKKKKK K
   (K){20}
47:  KKKKK KKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAB23585  ck: 9893  len: 36  ! Aab23585 Ask21 linker peptide. 1/2001
   (R,K){20}
15:  PGSGS KKKKKKKKKKKKKKKKKKKKK KG
   (K){20}
16:  GSGSK KKKKKKKKKKKKKKKKKKKKK G
   (K){20}

AAB23586  ck: 58  len: 36  ! Aab23586 Gsk21 linker peptide. 1/2001
   (R,K){20}
15:  SGSGS KKKKKKKKKKKKKKKKKKKKK KG
   (K){20}
16:  GSGSK KKKKKKKKKKKKKKKKKKKKK G
   (K){20}

AAB23591  ck: 7684  len: 630  ! Aab23591 Modified fibre protein encoded in
   (R,K){20}

```

```

596:  PGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
   (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
   (K){20}

AAB23592  ck: 7647  len: 630  ! Aab23592 Modified fibre protein encoded
   (R,K){20}
596:  SGSGS KKKKKKKKKKKKKKKKKKKKK KGSYS
   (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSYSM
   (K){20}

AAB23593  ck: 2249  len: 640  ! Aab23593 Modified fibre protein encoded
   (R,K){20}
596:  PGSGS KKKKKKKKKKKKKKKKKKKKK KGSAE
   (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSAEK
   (K){20}

AAB23594  ck: 2612  len: 640  ! Aab23594 Modified fibre protein encoded
   (R,K){20}
596:  SGSGS KKKKKKKKKKKKKKKKKKKKK KGSAE
   (K){20}
597:  GSGSK KKKKKKKKKKKKKKKKKKKKK GSAEK
   (K){20}

AAB13780  ck: 7317  len: 21  ! Aab13780 Soluble peptide antigen pk. 11/
   (R,K){20}
2:  C KKKKKKKKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAB13783  ck: 4553  len: 45  ! Aab13783 Soluble tandem pEA/ PK peptide
   (R,K){20}
26:  AAAAA KKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAB13784  ck: 4126  len: 44  ! Aab13784 Soluble tandem HA/ PK peptide c
   (R,K){20}
25:  DGWYG KKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAG00834  ck: 6330  len: 103  ! Aag00834 Human secreted protein, SEQ ID
   (R,K){20}
83:  RKROQ KKKKKKKKKKKKKKKKKKKKK K
   (K){20}
84:  KRQOK KKKKKKKKKKKKKKKKKKKKK
   (K){20}

AAY98493  ck: 8137  len: 45  ! Aay98493 Peptide #5 used in nucleic acid
   (R,K){20}
4:  YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
   (K){20}

```

5: YKAK (K){20} KKKKKK
6: YKAK (K){20} KKKKKK
7: KAKKK (K){20} KKKKKK
8: AKKKK (K){20} KKKKKK
9: KKKKK (K){20} KKKKKK
10: KKKKK (K){20} KKKKKK
11: KKKKK (K){20} KKKKKK
12: KKKKK (K){20} KKKKKK
13: KKKKK (K){20} KKKKKK
14: KKKKK (K){20} KKKKKK
15: KKKKK (K){20} KKKKKK
16: KKKKK (K){20} KKKKKK
17: KKKKK (K){20} KKKKKK
18: KKKKK (K){20} KKKKKK
19: KKKKK (K){20} KKKKKK
20: KKKKK (K){20} KKKKKK
21: KKKKK (K){20} KKKKKK
22: KKKKK (K){20} KKKKKK
23: KKKKK (K){20} KKKKKK
24: KKKKK (K){20} KKKKKK

1 AAY98495 ck: 4361 len: 59 1 Aay98495 Nuclear ligand used in nucleic aci

18: ADYKA (R,K){20} KKKKKK
19: PYKAK (K){20} KKKKKK
20: YKAKK (K){20} KKKKKK

21: KAKKK (K){20} KKKKKK
22: AKKKK (K){20} KKKKKK
23: KKKKK (K){20} KKKKKK
24: KKKKK (K){20} KKKKKK
25: KKKKK (K){20} KKKKKK
26: KKKKK (K){20} KKKKKK
27: KKKKK (K){20} KKKKKK
28: KKKKK (K){20} KKKKKK
29: KKKKK (K){20} KKKKKK
30: KKKKK (K){20} KKKKKK
31: KKKKK (K){20} KKKKKK
32: KKKKK (K){20} KKKKKK
33: KKKKK (K){20} KKKKKK
34: KKKKK (K){20} KKKKKK
35: KKKKK (K){20} KKKKKK
36: KKKKK (K){20} KKKKKK
37: KKKKK (K){20} KKKKKK
38: KKKKK (K){20} KKKKKK

1 AAY98497 ck: 4925 len: 100 1 Aay98497 Peptide #8 used in nucleic aci

1: KKKKKKKKKKKKKKKKKKK KKKKKK
2: K (K){20} KKKKKK
3: KK (K){20} KKKKKK
4: KKK (K){20} KKKKKK
5: KKKK (K){20} KKKKKK

6: KKKK (K){20} KKKK
 7: KKKK (K){20} KKKK
 8: KKKK (K){20} KKKK
 9: KKKK (K){20} KKKK
 10: KKKK (K){20} KKKK
 11: KKKK (K){20} KKKK
 12: KKKK (K){20} KKKK
 13: KKKK (K){20} KKKK
 14: KKKK (K){20} KKKK
 15: KKKK (K){20} KKKK
 16: KKKK (K){20} KKKK
 17: KKKK (K){20} KKKK
 18: KKKK (K){20} KKKK
 19: KKKK (K){20} KKKK
 20: KKKK (K){20} KKKK
 21: KKKK (K){20} KKKK
 22: KKKK (K){20} KKKK
 23: KKKK (K){20} KKKK
 24: KKKK (K){20} KKKK
 25: KKKK (K){20} KKKK
 26: KKKK (K){20} KKKK
 27: KKKK (K){20} KKKK
 28: KKKK (K){20} KKKK
 29: KKKK (K){20} KKKK

30: KKKK (K){20} KKKK
 31: KKKK (K){20} KKKK
 32: KKKK (K){20} KKKK
 33: KKKK (K){20} KKKK
 34: KKKK (K){20} KKKK
 35: KKKK (K){20} KKKK
 36: KKKK (K){20} KKKK
 37: KKKK (K){20} KKKK
 38: KKKK (K){20} KKKK
 39: KKKK (K){20} KKKK
 40: KKKK (K){20} KKKK
 41: KKKK (K){20} KKKK
 42: KKKK (K){20} KKKK
 43: KKKK (K){20} KKKK
 44: KKKK (K){20} KKKK
 45: KKKK (K){20} KKKK
 46: KKKK (K){20} KKKK
 47: KKKK (K){20} KKKK
 48: KKKK (K){20} KKKK
 49: KKKK (K){20} KKKK
 50: KKKK (K){20} KKKK
 51: KKKK (K){20} KKKK
 52: KKKK (K){20} KKKK
 53: KKKK (K){20} KKKK

54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 67: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 68: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 69: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 70: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 71: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 72: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 73: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 74: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 75: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 76: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 77: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 78: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

79: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 81: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

1
 AAY81805 ck: 1885 len: 351 1 Aay81805 Murine mahogany protein sequen
 (R,K){20}
 (R,K){20}
 3: EF KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 4: EFR KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 5: EFRK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 6: EFRKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 7: FRKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 8: RKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
25: KKKK (K){20} KKKK
26: KKKK (K){20} KKKK
27: KKKK (K){20} KKKK

AAY56902 ck: 4875 len: 30 1 Aay56902 (Lys)30 protein sequence. 4/2000

1

1: - (R,K){20} KKKK
2: K (K){20} KKKK
3: KK (K){20} KKKK
4: KKK (K){20} KKKK
5: KKKK (K){20} KKKK
6: KKKK (K){20} KKKK
7: KKKK (K){20} KKKK
8: KKKK (K){20} KKKK
9: KKKK (K){20} KKKK
10: KKKK (K){20} KKKK
11: KKKK (K){20} KKKK

AAY56903 ck: 5075 len: 434 1 Aay56903 (Lys)434 protein sequence. 4/2000

1

1: (R,K){20} KKKK
2: K (K){20} KKKK
3: KK (K){20} KKKK
4: KKK (K){20} KKKK
5: KKKK (K){20} KKKK

6: KKKK (K){20} KKKK
7: KKKK (K){20} KKKK
8: KKKK (K){20} KKKK
9: KKKK (K){20} KKKK
10: KKKK (K){20} KKKK
11: KKKK (K){20} KKKK
12: KKKK (K){20} KKKK
13: KKKK (K){20} KKKK
14: KKKK (K){20} KKKK
15: KKKK (K){20} KKKK
16: KKKK (K){20} KKKK
17: KKKK (K){20} KKKK
18: KKKK (K){20} KKKK
19: KKKK (K){20} KKKK
20: KKKK (K){20} KKKK
21: KKKK (K){20} KKKK
22: KKKK (K){20} KKKK
23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
25: KKKK (K){20} KKKK
26: KKKK (K){20} KKKK
27: KKKK (K){20} KKKK
28: KKKK (K){20} KKKK
29: KKKK (K){20} KKKK
30: KKKK (K){20} KKKK

31: KKKK (K){20} KKKK
32: KKKK (K){20} KKKK
33: KKKK (K){20} KKKK
34: KKKK (K){20} KKKK
35: KKKK (K){20} KKKK
36: KKKK (K){20} KKKK
37: KKKK (K){20} KKKK
38: KKKK (K){20} KKKK
39: KKKK (K){20} KKKK
40: KKKK (K){20} KKKK
41: KKKK (K){20} KKKK
42: KKKK (K){20} KKKK
43: KKKK (K){20} KKKK
44: KKKK (K){20} KKKK
45: KKKK (K){20} KKKK
46: KKKK (K){20} KKKK
47: KKKK (K){20} KKKK
48: KKKK (K){20} KKKK
49: KKKK (K){20} KKKK
50: KKKK (K){20} KKKK
51: KKKK (K){20} KKKK
52: KKKK (K){20} KKKK
53: KKKK (K){20} KKKK
54: KKKK (K){20} KKKK

55: KKKK (K){20} KKKK
56: KKKK (K){20} KKKK
57: KKKK (K){20} KKKK
58: KKKK (K){20} KKKK
59: KKKK (K){20} KKKK
60: KKKK (K){20} KKKK
61: KKKK (K){20} KKKK
62: KKKK (K){20} KKKK
63: KKKK (K){20} KKKK
64: KKKK (K){20} KKKK
65: KKKK (K){20} KKKK
66: KKKK (K){20} KKKK
67: KKKK (K){20} KKKK
68: KKKK (K){20} KKKK
69: KKKK (K){20} KKKK
70: KKKK (K){20} KKKK
71: KKKK (K){20} KKKK
72: KKKK (K){20} KKKK
73: KKKK (K){20} KKKK
74: KKKK (K){20} KKKK
75: KKKK (K){20} KKKK
76: KKKK (K){20} KKKK
77: KKKK (K){20} KKKK
78: KKKK (K){20} KKKK

79: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
80: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
81: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
82: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
83: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
84: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
85: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
86: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
87: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
88: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
89: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
90: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
91: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
92: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
93: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
94: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
95: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
96: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
97: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
98: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
99: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
100: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
101: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
102: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
103: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

104: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
105: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
106: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
107: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
108: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
109: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
110: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
111: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
112: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
113: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
114: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
115: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
116: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
117: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
118: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
119: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
120: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
121: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
122: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
123: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
124: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
125: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
126: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
127: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

128: KKKK (K){20} KKKK
129: KKKK (K){20} KKKK
130: KKKK (K){20} KKKK
131: KKKK (K){20} KKKK
132: KKKK (K){20} KKKK
133: KKKK (K){20} KKKK
134: KKKK (K){20} KKKK
135: KKKK (K){20} KKKK
136: KKKK (K){20} KKKK
137: KKKK (K){20} KKKK
138: KKKK (K){20} KKKK
139: KKKK (K){20} KKKK
140: KKKK (K){20} KKKK
141: KKKK (K){20} KKKK
142: KKKK (K){20} KKKK
143: KKKK (K){20} KKKK
144: KKKK (K){20} KKKK
145: KKKK (K){20} KKKK
146: KKKK (K){20} KKKK
147: KKKK (K){20} KKKK
148: KKKK (K){20} KKKK
149: KKKK (K){20} KKKK
150: KKKK (K){20} KKKK
151: KKKK (K){20} KKKK

152: KKKK KKKKKKKKKKKKKKKKKKK KKKK
153: KKKK (K){20} KKKK
154: KKKK (K){20} KKKK
155: KKKK (K){20} KKKK
156: KKKK (K){20} KKKK
157: KKKK (K){20} KKKK
158: KKKK (K){20} KKKK
159: KKKK (K){20} KKKK
160: KKKK (K){20} KKKK
161: KKKK (K){20} KKKK
162: KKKK (K){20} KKKK
163: KKKK (K){20} KKKK
164: KKKK (K){20} KKKK
165: KKKK (K){20} KKKK
166: KKKK (K){20} KKKK
167: KKKK (K){20} KKKK
168: KKKK (K){20} KKKK
169: KKKK (K){20} KKKK
170: KKKK (K){20} KKKK
171: KKKK (K){20} KKKK
172: KKKK (K){20} KKKK
173: KKKK (K){20} KKKK
174: KKKK (K){20} KKKK
175: KKKK (K){20} KKKK
176: KKKK (K){20} KKKK

177: KKKKK (K){20} KKKKK
 178: KKKKK (K){20} KKKKK
 179: KKKKK (K){20} KKKKK
 180: KKKKK (K){20} KKKKK
 181: KKKKK (K){20} KKKKK
 182: KKKKK (K){20} KKKKK
 183: KKKKK (K){20} KKKKK
 184: KKKKK (K){20} KKKKK
 185: KKKKK (K){20} KKKKK
 186: KKKKK (K){20} KKKKK
 187: KKKKK (K){20} KKKKK
 188: KKKKK (K){20} KKKKK
 189: KKKKK (K){20} KKKKK
 190: KKKKK (K){20} KKKKK
 191: KKKKK (K){20} KKKKK
 192: KKKKK (K){20} KKKKK
 193: KKKKK (K){20} KKKKK
 194: KKKKK (K){20} KKKKK
 195: KKKKK (K){20} KKKKK
 196: KKKKK (K){20} KKKKK
 197: KKKKK (K){20} KKKKK
 198: KKKKK (K){20} KKKKK
 199: KKKKK (K){20} KKKKK
 200: KKKKK (K){20} KKKKK



201: KKKKK (K){20} KKKKK
 202: KKKKK (K){20} KKKKK
 203: KKKKK (K){20} KKKKK
 204: KKKKK (K){20} KKKKK
 205: KKKKK (K){20} KKKKK
 206: KKKKK (K){20} KKKKK
 207: KKKKK (K){20} KKKKK
 208: KKKKK (K){20} KKKKK
 209: KKKKK (K){20} KKKKK
 210: KKKKK (K){20} KKKKK
 211: KKKKK (K){20} KKKKK
 212: KKKKK (K){20} KKKKK
 213: KKKKK (K){20} KKKKK
 214: KKKKK (K){20} KKKKK
 215: KKKKK (K){20} KKKKK
 216: KKKKK (K){20} KKKKK
 217: KKKKK (K){20} KKKKK
 218: KKKKK (K){20} KKKKK
 219: KKKKK (K){20} KKKKK
 220: KKKKK (K){20} KKKKK
 221: KKKKK (K){20} KKKKK
 222: KKKKK (K){20} KKKKK
 223: KKKKK (K){20} KKKKK
 224: KKKKK (K){20} KKKKK

225: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
226: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
227: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
228: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
229: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
230: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
231: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
232: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
233: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
234: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
235: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
236: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
237: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
238: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
239: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
240: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
241: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
242: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
243: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
244: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
245: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
246: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
247: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
248: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
249: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}

250: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
251: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
252: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
253: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
254: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
255: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
256: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
257: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
258: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
259: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
260: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
261: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
262: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
263: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
264: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
265: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
266: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
267: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
268: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
269: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
270: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
271: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
272: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
273: KKKK KKKKKKKKKKKKKKKKKKK KKKK
 (K){20}

274: KKKK (K)[20]
 275: KKKK (K)[20]
 276: KKKK (K)[20]
 277: KKKK (K)[20]
 278: KKKK (K)[20]
 279: KKKK (K)[20]
 280: KKKK (K)[20]
 281: KKKK (K)[20]
 282: KKKK (K)[20]
 283: KKKK (K)[20]
 284: KKKK (K)[20]
 285: KKKK (K)[20]
 286: KKKK (K)[20]
 287: KKKK (K)[20]
 288: KKKK (K)[20]
 289: KKKK (K)[20]
 290: KKKK (K)[20]
 291: KKKK (K)[20]
 292: KKKK (K)[20]
 293: KKKK (K)[20]
 294: KKKK (K)[20]
 295: KKKK (K)[20]
 296: KKKK (K)[20]
 297: KKKK (K)[20]

298: KKKK
 299: KKKK (K)[20]
 300: KKKK (K)[20]
 301: KKKK (K)[20]
 302: KKKK (K)[20]
 303: KKKK (K)[20]
 304: KKKK (K)[20]
 305: KKKK (K)[20]
 306: KKKK (K)[20]
 307: KKKK (K)[20]
 308: KKKK (K)[20]
 309: KKKK (K)[20]
 310: KKKK (K)[20]
 311: KKKK (K)[20]
 312: KKKK (K)[20]
 313: KKKK (K)[20]
 314: KKKK (K)[20]
 315: KKKK (K)[20]
 316: KKKK (K)[20]
 317: KKKK (K)[20]
 318: KKKK (K)[20]
 319: KKKK (K)[20]
 320: KKKK (K)[20]
 321: KKKK (K)[20]
 322: KKKK (K)[20]

323: KKKK (K) [20] KKKK
324: KKKK (K) [20] KKKK
325: KKKK (K) [20] KKKK
326: KKKK (K) [20] KKKK
327: KKKK (K) [20] KKKK
328: KKKK (K) [20] KKKK
329: KKKK (K) [20] KKKK
330: KKKK (K) [20] KKKK
331: KKKK (K) [20] KKKK
332: KKKK (K) [20] KKKK
333: KKKK (K) [20] KKKK
334: KKKK (K) [20] KKKK
335: KKKK (K) [20] KKKK
336: KKKK (K) [20] KKKK
337: KKKK (K) [20] KKKK
338: KKKK (K) [20] KKKK
339: KKKK (K) [20] KKKK
340: KKKK (K) [20] KKKK
341: KKKK (K) [20] KKKK
342: KKKK (K) [20] KKKK
343: KKKK (K) [20] KKKK
344: KKKK (K) [20] KKKK
345: KKKK (K) [20] KKKK
346: KKKK (K) [20] KKKK

347: KKKK (K) [20] KKKK
348: KKKK (K) [20] KKKK
349: KKKK (K) [20] KKKK
350: KKKK (K) [20] KKKK
351: KKKK (K) [20] KKKK
352: KKKK (K) [20] KKKK
353: KKKK (K) [20] KKKK
354: KKKK (K) [20] KKKK
355: KKKK (K) [20] KKKK
356: KKKK (K) [20] KKKK
357: KKKK (K) [20] KKKK
358: KKKK (K) [20] KKKK
359: KKKK (K) [20] KKKK
360: KKKK (K) [20] KKKK
361: KKKK (K) [20] KKKK
362: KKKK (K) [20] KKKK
363: KKKK (K) [20] KKKK
364: KKKK (K) [20] KKKK
365: KKKK (K) [20] KKKK
366: KKKK (K) [20] KKKK
367: KKKK (K) [20] KKKK
368: KKKK (K) [20] KKKK
369: KKKK (K) [20] KKKK
370: KKKK (K) [20] KKKK

371: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 372: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 373: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 374: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 375: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 376: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 377: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 378: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 379: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 380: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 381: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 382: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 383: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 384: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 385: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 386: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 387: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 388: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 389: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 390: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 391: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 392: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 393: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 394: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 395: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

396: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 397: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 398: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 399: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 400: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 401: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 402: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 403: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 404: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 405: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 406: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 407: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 408: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 409: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 410: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 411: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 412: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 413: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 414: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 415: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

1

AY86248 ck: 8783 len: 128 1 Aay86248 Human secreted protein HCHPF68,
 (R,K){20}
 84: ANPP KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 85: NPSP KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 86: PPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}

87: PPKK (K){20} KKKK
88: PKKK (K){20} KKKK
89: KKKK (K){20} KKKK
90: KKKK (K){20} KKKK
91: KKKK (K){20} KKKK
92: KKKK (K){20} KKKK
93: KKKK (K){20} KKKK
94: KKKK (K){20} KKKK
95: KKKK (K){20} KKKK
96: KKKK (K){20} KKKK
97: KKKK (K){20} KKKK
98: KKKK (K){20} KKKK
99: KKKK (K){20} KKKK
100: KKKK (K){20} KKKK
101: KKKK (K){20} KKKK
102: KKKK (K){20} KKKK
103: KKKK (K){20} KKKK
104: KKKK (K){20} KKKK
105: KKKK (K){20} KKKK
106: KKKK (K){20} KKKK
AAV59038 ck: 8137 len: 45 1 Aay59038 Peptide used in the construction c

1
4: YKA (R,K){20} KKKK
5: YKA (K){20} KKKK
6: YKA (K){20} KKKK

7: KAKK (K){20} KKKK
8: AKKK (K){20} KKKK
9: KKKK (K){20} KKKK
10: KKKK (K){20} KKKK
11: KKKK (K){20} KKKK
12: KKKK (K){20} KKKK
13: KKKK (K){20} KKKK
14: KKKK (K){20} KKKK
15: KKKK (K){20} KKKK
16: KKKK (K){20} KKKK
17: KKKK (K){20} KKKK
18: KKKK (K){20} KKKK
19: KKKK (K){20} KKKK
20: KKKK (K){20} KKKK
21: KKKK (K){20} KKKK
22: KKKK (K){20} KKKK
23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
AAV59040 ck: 4361 len: 59 1 Aay59040 Nuclear ligand for transportin

1
18: APYA (R,K){20} KKKK
19: PYAK (K){20} KKKK
20: YAKK (K){20} KKKK
21: KAKK (K){20} KKKK
22: AKKK (K){20} KKKK

23: KKKK (K){20} KKKK
 24: KKKK (K){20} KKKK
 25: KKKK (K){20} KKKK
 26: KKKK (K){20} KKKK
 27: KKKK (K){20} KKKK
 28: KKKK (K){20} KKKK
 29: KKKK (K){20} KKKK
 30: KKKK (K){20} KKKK
 31: KKKK (K){20} KKKK
 32: KKKK (K){20} KKKK
 33: KKKK (K){20} KKKK
 34: KKKK (K){20} KKKK
 35: KKKK (K){20} KKKK
 36: KKKK (K){20} KKKK
 37: KKKK (K){20} KKKK
 38: KKKK (K){20} KKKK

AA59042 ck: 4925 len: 100 1 Aay59042 Amino acid polymer seg ID NO: 62 C

1: (R,K){20} KKKK
 2: K (K){20} KKKK
 3: KK (K){20} KKKK
 4: KKK (K){20} KKKK
 5: KKKK (K){20} KKKK
 6: KKKK (K){20} KKKK
 7: KKKK (K){20} KKKK

8: KKKK (K){20} KKKK
 9: KKKK (K){20} KKKK
 10: KKKK (K){20} KKKK
 11: KKKK (K){20} KKKK
 12: KKKK (K){20} KKKK
 13: KKKK (K){20} KKKK
 14: KKKK (K){20} KKKK
 15: KKKK (K){20} KKKK
 16: KKKK (K){20} KKKK
 17: KKKK (K){20} KKKK
 18: KKKK (K){20} KKKK
 19: KKKK (K){20} KKKK
 20: KKKK (K){20} KKKK
 21: KKKK (K){20} KKKK
 22: KKKK (K){20} KKKK
 23: KKKK (K){20} KKKK
 24: KKKK (K){20} KKKK
 25: KKKK (K){20} KKKK
 26: KKKK (K){20} KKKK
 27: KKKK (K){20} KKKK
 28: KKKK (K){20} KKKK
 29: KKKK (K){20} KKKK
 30: KKKK (K){20} KKKK
 31: KKKK (K){20} KKKK

32: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
33: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
34: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
35: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
36: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
37: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
38: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
39: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
40: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
41: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
42: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
43: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
44: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
45: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
46: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
47: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
48: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
49: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
50: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
51: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
52: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
53: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
54: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
55: KKKK (K) [20]
KKKKKKKKKKKKKKKKKK KKKK
(K) [20]

56: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
57: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
58: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
59: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
60: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
61: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
62: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
63: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
64: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
65: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
66: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
67: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
68: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
69: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
70: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
71: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
72: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
73: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
74: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
75: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
76: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
77: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
78: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
79: KKKK KKKKKKKKKKKKKKKKK KKKK
(K) [20]
80: KKKK KKKKKKKKKKKKKKKKK K

81: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK

ABG00401 ck: 5116 len: 1,074 i Abg00401 Novel human diagnostic protein #35

609: RGSSS (R,K){20}
KKKKRRKKKKKKKKKKRR KNRKK
(R,K){20}

610: GSSSK (R,K){20}
KKRRKKKKKKKKRRRRR NRKKK

ABG03974 ck: 5341 len: 99 i Abg03974 Novel human diagnostic protein #35

1 (R,K){20}

2: M KKKKKKKKKKKKKKKKK KKKKN
(K){20}

3: MK KKKKKKKKKKKKKKKKK KKKNS
(K){20}

4: MKK KKKKKKKKKKKKKKKKK KKNSS
(K){20}

5: MKKK KKKKKKKKKKKKKKKKK KNSQI
(K){20}

6: MKKKK KKKKKKKKKKKKKKKKK NSQID
(K){20}

ABG04391 ck: 9047 len: 139 i Abg04391 Novel human diagnostic protein #43

1 (R,K){20}

79: EEEEE (R,K){20}
KKKKKKKKKKKKKKKKKK KKKKK
(K){20}

80: EEEKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

81: EEEKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

82: EEEKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

83: EKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

84: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

85: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

86: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

87: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

88: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

89: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

90: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

91: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

92: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

93: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

94: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

95: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

96: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

97: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

98: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

99: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

100: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

101: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

102: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

103: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

104: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

105: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

106: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

107: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

108: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

109: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKE

110: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKEN

111: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKENP

112: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KENPF

113: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK ENPFS

ABG05352 ck: 1276 len: 204 i Abg05352 Novel human diagnostic protein

106: EEEEE (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

107: EEEEE (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

108: EEEEE (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

109: EEEEE (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

110: EEEEE (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

111: KKKKK (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

112: KKKKK (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

113: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

114: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

115: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

116: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

117: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

118: KKKKK (R,K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

ABG05367 ck: 6907 len: 808 1 Abg05367 Novel human diagnostic protein #53

219: EEEEE (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

220: EEEEE (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

221: EEEEE (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

222: EEEEE (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

223: EEEEE (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

224: RKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

225: RKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

226: KKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

227: KKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

228: KKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

1

229: KKKKK (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

230: KKKKK (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

231: KKKKK (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

232: KKKKK (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

233: KKKKK (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

234: RRRRR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

235: RRRRR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

236: RRRRR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

237: RRRRR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

238: KKKKK (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

239: KKKKK (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

240: KKKKK (K) [20]
KKKKKKKKKKKKKKKKKKKKKKKKKK

ABG06375 ck: 7807 len: 2,570 1 Abg06375 Novel human diagnostic protein

205: KKKSS (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

206: KSSSR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

207: KSSSR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

208: SSRRR (R,K) [20]
RRRRRRRRRRRRRRRRRRRRRRRRRR

ABG06513 ck: 2934 len: 154 1 Abg06513 Novel human diagnostic protein

49: KKKKT (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

50: KKKTR (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

51: KKTTR (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

52: KTRKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

53: TRKKR (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

54: RKKKK (R,K) [20]
RRKKKKKKKKKKKKKKKKKKKKKKKK

55: KKKKK (R,K){20} KKKKKK
56: KKKKK (R,K){20} KKKKKK
57: KKKKK (R,K){20} KKKKKK
58: KKKKK (R,K){20} KKKKKK
59: KKKKK (R,K){20} KKKKKK
60: KKKKK (R,K){20} KKKKKK
61: KKKKK (R,K){20} KKKKKK
62: KKKKK (R,K){20} KKKKKK
63: KKKKK (R,K){20} KKKKKK
64: KKKKK (R,K){20} KKKKKK
65: KKKKK (R,K){20} KKKKKK
ABG07742 ck: 8672 len: 502 ! Abg07742 Novel human diagnostic protein #77
54: DDEEE (R,K){20} RRRRRR
55: DEEER (R,K){20} RRRRRR
56: EEEER (R,K){20} RRRRRR
57: EERRR (R,K){20} RRRRRR
58: ERRRR (R,K){20} RRRRRR
59: RRRRR (R,K){20} KKKKKK
60: RRRRR (R,K){20} KKKKKK
61: RRRRR (R,K){20} KKKKKK
62: KKKKK (R,K){20} KKKKKK
63: KKKKK (R,K){20} KKKKKK
64: KKKKK (R,K){20} KKKKKK
65: KKKKK (R,K){20} KKKKKK
ABG11241 ck: 3870 len: 121 ! Abg11241 Novel human diagnostic protein #77
54: DDEEE (R,K){20} RRRRRR
55: DEEER (R,K){20} RRRRRR
56: EEEER (R,K){20} RRRRRR
57: EERRR (R,K){20} RRRRRR
58: ERRRR (R,K){20} RRRRRR
59: RRRRR (R,K){20} KKKKKK
60: RRRRR (R,K){20} KKKKKK
61: RRRRR (R,K){20} KKKKKK
62: KKKKK (R,K){20} KKKKKK
63: KKKKK (R,K){20} KKKKKK
64: KKKKK (R,K){20} KKKKKK
65: KKKKK (R,K){20} KKKKKK

66: KKKKK (R,K){20} KKKKKK
67: KKKKK (R,K){20} KKKKKK
ABG10052 ck: 7107 len: 39 ! Abg10052 Novel human diagnostic protein
2: Q KKKKK (R,K){20} KKKKKK
3: Q KKKKK (R,K){20} KKKKKK
4: Q KKKKK (R,K){20} KKKKKK
5: Q KKKKK (R,K){20} KKKKKK
6: Q KKKKK (R,K){20} KKKKKK
7: KKKKK (R,K){20} KKKKKK
ABG10053 ck: 3274 len: 189 ! Abg10053 Novel human diagnostic protein
70: EKEKE (R,K){20} KKKKKK
71: KEKEK (R,K){20} KKKKKK
72: EKEKE (R,K){20} KKKKKK
73: KEKEK (R,K){20} KKKKKK
74: EKEKE (R,K){20} KKKKKK
75: KEKEK (R,K){20} KKKKKK
76: EKEKE (R,K){20} KKKKKK
77: KEKEK (R,K){20} KKKKKK
78: EKEKE (R,K){20} KKKKKK
79: KEKEK (R,K){20} KKKKKK
ABG11241 ck: 3870 len: 121 ! Abg11241 Novel human diagnostic protein
44: KKKKE (R,K){20} KKKKKK
45: KKKKE (R,K){20} KKKKKK

46: KEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE
47: EEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
48: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKE
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG11242 ck: 4493 len: 100 ! Abg11242 Novel human diagnostic protein #11

1
(R,K){20}

42: ETPSE (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

43: TPSEK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

44: PSEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

45: SEKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

46: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

58: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

1

61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

63: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

64: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

65: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

66: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

67: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

68: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

69: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

70: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

71: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

ABG11245 ck: 2517 len: 85 ! Abg11245 Novel human diagnostic protein

(R,K){20}

33: EEEBE (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

34: EEEBK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

35: EEEBK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

36: EEEKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

37: EKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

45: KKKKK (K){20} KKKKK
46: KKKKK (K){20} KKKKK
47: KKKKK (K){20} KKKKK
48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
50: KKKKK (K){20} KKKKK
51: KKKKK (K){20} KKKKK
52: KKKKK (K){20} KKKKK
ABG11250 ck: 4343 len: 92 ! Abg11250 Novel human diagnostic protein #11
(R,K){20}
53: EKEKE (R,K){20} KKKKK
(K){20}
54: KEKEK (K){20} KKKKK
(K){20}
55: EKEKK (K){20} KKKKK
(K){20}
56: KEKKK (K){20} KKKKK
(R,K){20}
57: EKKKK (R,K){20} KKKKK
(R,K){20}
58: KKKKK (R,K){20} KKKKK
(R,K){20}
59: KKKKK (R,K){20} KKKKK
ABG11266 ck: 9563 len: 146 ! Abg11266 Novel human diagnostic protein #11
(R,K){20}
36: KEKKE (K){20} KKKKK
(K){20}
37: EKREK (K){20} KKKKK
(K){20}
38: KEKKK (K){20} KKKKK
(K){20}
39: REKKK (K){20} KKKKK
(K){20}
40: EKKKK (K){20} EEEEE
(R,K){20}
79: EEEED (R,K){20} KKKKK
(R,K){20}
80: EEEEE (R,K){20} KKKKK

1
81: EEEKK (R,K){20} KKKKK
(R,K){20}
82: EEEKK (R,K){20} KKKKK
(R,K){20}
83: EKKKK (R,K){20} KKKKK
(R,K){20}
84: KKKKK (R,K){20} KKKKK
(R,K){20}
85: KKKKK (R,K){20} KKKKK
(R,K){20}
86: KKKKK (R,K){20} KKKKK
(R,K){20}
87: KKKKK (R,K){20} KKKKK
(R,K){20}
88: KKKKK (R,K){20} KKKKK
(R,K){20}
89: KKKKK (R,K){20} KKKKK
(R,K){20}
90: KKKKK (R,K){20} KKKKK
(R,K){20}
91: KKKKK (R,K){20} KKKKK
(R,K){20}
92: KKKKK (R,K){20} KKKKK
(R,K){20}
93: KKKKK (R,K){20} KKKKK
ABG11277 ck: 8026 len: 1,080 ! Abg11277 Novel human diagnostic protein
(R,K){20}
709: QKEKE (R,K){20} KKKKK
(R,K){20}
710: KEKEK (R,K){20} KKKKK
(R,K){20}
711: EKEKK (R,K){20} KKKKK
(R,K){20}
712: KEKKK (R,K){20} KKKKK
(R,K){20}
713: EKKKK (R,K){20} KKKKK
(R,K){20}
714: KKKKK (R,K){20} KKKKK
(R,K){20}
715: KKKKK (R,K){20} KKKKK
(R,K){20}
716: KKKKK (R,K){20} KKKKK
(R,K){20}
717: KKKKK (R,K){20} KKKKK
(R,K){20}
718: KKKKK (R,K){20} KKKKK

```

719: KKKKK (R,K){20}
720: KKKKK (R,K){20}
721: KKKKK (R,K){20}
722: KKKKK (R,K){20}
723: KKKKK (R,K){20}
724: KKKKK (R,K){20}
725: KKKKK (R,K){20}
726: KKKKK (R,K){20}
727: KKKKK (R,K){20}

ABG11732 ck: 2886 len: 56 1 Abg11732 Novel human diagnostic protein #11
8: KRRRG RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
9: RRRGR RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
10: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
11: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
12: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
13: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}

ABG11734 ck: 4548 len: 142 1 Abg11734 Novel human diagnostic protein #11
109: QMLSV KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
110: MLSVK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
111: LSVKK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
112: SVKKK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
113: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
      (K){20}
```

```

115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKLY (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKLVF (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KLYFQ (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK LYFQT (K){20}

ABG11738 ck: 3009 len: 567 1 Abg11738 Novel human diagnostic protein
526: EEEED RKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
527: EEEED RKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
528: EEEER KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
529: EERRR KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
530: ERRKK KKKKKKKKKKKKKKKKKKKK KKKKE (K){20}
531: RRRKK KKKKKKKKKKKKKKKKKKKK KKKER (K){20}
532: RKKKK KKKKKKKKKKKKKKKKKKKK KKKER (K){20}
533: KKKKK KKKKKKKKKKKKKKKKKKKK KKKERT (R,K){20}
534: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRTA (R,K){20}

ABG22512 ck: 8641 len: 856 1 Abg22512 Novel human diagnostic protein
130: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (R,K){20}
131: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
132: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
133: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
134: FKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
135: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}
136: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK (K){20}

ABG22638 ck: 4197 len: 896 1 Abg22638 Novel human diagnostic protein
227: GREER RRRRRRRRRRRRRRRRRRRR RRRRR (R,K){20}
      (R,K){20}
```

228: RERER RRRRRRRRRRRRRRRRRR RRRGX (R,K){20}
229: ERERR RRRRRRRRRRRRRRRRRR RRGXE (R,K){20}
230: RERRR RRRRRRRRRRRRRRRRRR RGXEF (R,K){20}
231: EERRR RRRRRRRRRRRRRRRRRR GXEFL (R,K){20}

ABG26213 ck: 6773 len: 735 1 Abg26213 Novel human diagnostic protein #26
(R,K){20}
(R,K){20}
173: RGSSS KKKRRRRRRRRRRRRRRR KNRRK
174: GSSSS KKKRRRRRRRRRRRRRRR NRKKK (R,K){20}

ABG26488 ck: 523 len: 124 1 Abg26488 Novel human diagnostic protein #26
(R,K){20}
(R,K){20}
91: EEEEE RRRRRRRRRRRRRRRRRR RKKKK
92: EEEER RRRRRRRRRRRRRRRRRR KKKKK (R,K){20}
93: EEERR RRRRRRRRRRRRRRRRRR KKKKR (R,K){20}
94: EERRR RRRRRRRRRRRRRRRRRR KKKRK (R,K){20}
95: ERRRR RRRRRRRRRRRRRRRRRR KKKKK (R,K){20}
96: RRRRR KKKRRRRRRRRRRRRRRR KRRKK (R,K){20}
97: RRRRK KKKRRRRRRRRRRRRRRR RKKKK (R,K){20}
98: RRRKK KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
99: RRRKK KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
100: RKKKK KKKRRRRRRRRRRRRRRR KKKKE (R,K){20}
101: KKKKK KKKRRRRRRRRRRRRRRR KKKK (R,K){20}
102: KKKKK KKKRRRRRRRRRRRRRRR KKK (R,K){20}
103: KKKKK KKKRRRRRRRRRRRRRRR KKK (R,K){20}
104: KKKKK KKKRRRRRRRRRRRRRRR E (R,K){20}

ABG26489 ck: 9923 len: 120 1 Abg26489 Novel human diagnostic protein #26
(R,K){20}

70: RRRRG RRRRRRRRRRRRRRRRRR RRRKK (R){20}
71: RRRGR RRRRRRRRRRRRRRRRRR RKKKE (R){20}
72: RGRGR RRRRRRRRRRRRRRRRRR KKKKE (R){20}
73: RGRRR RRRRRRRRRRRRRRRRRR KKEEE (R,K){20}
74: GRRRR RRRRRRRRRRRRRRRRRR KEEEE (R,K){20}
75: RRRRR RRRRRRRRRRRRRRRRRR EEEEE (R,K){20}

ABG26490 ck: 4146 len: 96 1 Abg26490 Novel human diagnostic protein
(R,K){20}
(R){20}
39: KKGEE RRRRRRRRRRRRRRRRRR RRRGR (R){20}
40: KGEE RRRRRRRRRRRRRRRRRR RRGGR (R){20}
41: GEEER RRRRRRRRRRRRRRRRRR RGRGR (R){20}
42: EERRR RRRRRRRRRRRRRRRRRR GRGGR (R){20}
73: GRGGR RRRRRRRRRRRRRRRRRR RRRR (R){20}
74: RGRGR RRRRRRRRRRRRRRRRRR RRR (R){20}
75: GRGGR RRRRRRRRRRRRRRRRRR RR (R){20}
76: RGRRR RRRRRRRRRRRRRRRRRR R (R){20}
77: GRRRR RRRRRRRRRRRRRRRRRR (R){20}

ABG26491 ck: 8179 len: 109 1 Abg26491 Novel human diagnostic protein
(R,K){20}
(R,K){20}
75: RRRGE KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
76: RRG EK KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
77: RGEKK KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
78: GEKKK KKKRRRRRRRRRRRRRRR KKKKK (R,K){20}
79: EKKKK KKKRRRRRRRRRRRRRRR KKKKG (R,K){20}
80: KKKKR KKKRRRRRRRRRRRRRRR KKKGN (R){20}
81: KKKRK KKKRRRRRRRRRRRRRRR KKGNT (R){20}

(K)(20)
82: KRRKK KKKKKKKKKKKKKKKK KGNLS
(K)(20)
83: KRRKK KKKKKKKKKKKKKKKK GNLSV
ABG26492 ck: 5234 len: 68 ! Abg26492 Novel human diagnostic protein #26
(R,K)(20)

29: KKKEE RRRRRRRRRRRRRRRRRR RRRRG
(R)(20)
30: KKEER RRRRRRRRRRRRRRRRRR RRRGK
(R)(20)
31: KEERR RRRRRRRRRRRRRRRRRR RRGKK
(R)(20)
32: EERRR RRRRRRRRRRRRRRRRRR RGKGD
(R)(20)
33: ERRRR RRRRRRRRRRRRRRRRRR GKKG

ABG26493 ck: 4204 len: 80 ! Abg26493 Novel human diagnostic protein #26

1
(R,K)(20)
27: EKEKE KRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
28: KEKEK RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
29: EKEKR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
30: KEKRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
31: EKRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
32: KRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
33: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
34: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
35: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
36: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
37: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
38: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
39: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
40: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)

41: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
42: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
43: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
44: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
45: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
46: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
47: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
48: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
49: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
50: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
51: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
52: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
ABG26496 ck: 8829 len: 90 ! Abg26496 Novel human diagnostic protein
(R,K)(20)
61: EKKEE KRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
62: KKEEK RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
63: KEEKR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
64: EEKRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
65: EKRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
66: KRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
67: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
68: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
69: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
70: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)

72: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
73: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
74: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
75: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
76: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
77: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
78: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
79: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
80: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
81: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
82: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
83: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
84: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
85: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
86: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
ABG26498 ck: 7156 len: 140 1 Abg26498 Novel human diagnostic protein #26
(R,K)(20)
97: RGGG RRRRRRRRRRRRRRRRRRR (R)(20)
98: RGGG RRRRRRRRRRRRRRRRRRR (R)(20)
99: GGGG RRRRRRRRRRRRRRRRRRR (R)(20)
100: GGGG RRRRRRRRRRRRRRRRRRR (R)(20)
101: GRRR RRRRRRRRRRRRRRRRRRR (R)(20)
102: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
103: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
104: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)

105: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
106: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
107: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
108: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
109: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
110: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
111: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
112: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
113: RRRR RRRRRRRRRRRRRRRRRRR (R,K)(20)
ABG26500 ck: 7252 len: 78 1 Abg26500 Novel human diagnostic protein
(R,K)(20)
35: GGEE RRRRRRRRRRRRRRRRRRR (R)(20)
36: GEEER RRRRRRRRRRRRRRRRRRR (R)(20)
37: EEER RRRRRRRRRRRRRRRRRRR (R)(20)
38: EERR RRRRRRRRRRRRRRRRRRR (R)(20)
39: EERRR RRRRRRRRRRRRRRRRRRR (R)(20)
40: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
41: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
42: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
43: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
44: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
45: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
46: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
47: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)
48: RRRR RRRRRRRRRRRRRRRRRRR (R)(20)

1

ABG26501 ck: 1730 len: 182 1 Abg26501 Novel human diagnostic protein #26

(R,K)(20)
(K)(20)
72: KEEKE KKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
73: EEEKE KKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
74: EEEKE KKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
75: KKKKK KKKKKKKKKKKKKKKKKKK KKKKE
(K)(20)
76: EKKKK KKKKKKKKKKKKKKKKKKK EEEEE
(R)(20)
133: GRRRS RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
134: RRRSR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
135: RRSRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
136: RSRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
137: SRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
138: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
139: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
140: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
141: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
142: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
143: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
144: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
145: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
146: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
147: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
148: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
149: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
150: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
YPMWV

1

ABG26502 ck: 3399 len: 101 1 Abg26502 Novel human diagnostic protein

(R,K)(20)
(R,K)(20)
59: EEEEG RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
60: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
61: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
62: EGRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
63: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
64: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
65: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
66: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
67: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
EYFQT

1

ABG26505 ck: 4704 len: 93 1 Abg26505 Novel human diagnostic protein

(R,K)(20)
(R,K)(20)
60: KEEEG RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
61: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
62: EEEGR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
63: EGRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
64: GRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
65: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R)(20)
66: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
67: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
68: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
69: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K)(20)
EEEE

1
ABG26506 ck: 3684 len: 85 ! Abg26506 Novel human diagnostic protein #26

(R,K){20}
47: EKEKE KKKKKKKKKKKKKKKKKKK

(R,K){20}
48: KEKEK KKKKKKKKKKKKKKKKKKK

(R,K){20}
49: EKEKE KKKKKKKKKKKKKKKKKKK

(R,K){20}
50: KEKEK KKKKKKKKKKKKKKKKKKK

(R,K){20}
51: EKRRK KKKKKKKKKKKKKKKKKKK

(R,K){20}
52: KRRKK KKKKKKKKKKKKKKKKKKK

(K){20}
53: RRRKK KKKKKKKKKKKKKKKKKKK

(K){20}
54: KRRKK KKKKKKKKKKKKKKKKKKK

(K){20}
55: RRRKK KKKKKKKKKKKKKKKKKKK

(K){20}
56: KRRKK KKKKKKKKKKKKKKKKKKK

(K){20}
57: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
58: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
59: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
60: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
61: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
62: KRRKK KKKKKKKKKKKKKKKKKKK

ABG26507 ck: 9838 len: 109 ! Abg26507 Novel human diagnostic protein #26

(R,K){20}
60: EEEEE RRRRRRRRRRRRRRRRRRR

(R){20}
61: EEEER RRRRRRRRRRRRRRRRRRR

(R){20}
62: EEEER RRRRRRRRRRRRRRRRRRR

(R){20}
63: EEEER RRRRRRRRRRRRRRRRRRR

(R,K){20}
64: EEEER RRRRRRRRRRRRRRRRRRR

(R,K){20}
65: EEEER RRRRRRRRRRRRRRRRRRR

1

ABG26508 ck: 4488 len: 121 ! Abg26508 Novel human diagnostic protein

(R,K){20}
56: DEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}
57: DEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}
58: DEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}
59: DEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}
60: EEEEE RRRRRRRRRRRRRRRRRRR

(R,K){20}
61: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
62: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
63: RRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
64: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
65: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
66: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
67: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
68: KRRKK KKKKKKKKKKKKKKKKKKK

(R,K){20}
69: KRRKK KKKKKKKKKKKKKKKKKKK

ABG26510 ck: 3426 len: 74 ! Abg26510 Novel human diagnostic protein

(R,K){20}
47: RRRRS RRRRRRRRRRRRRRRRRRR

(R){20}
48: RRRSR RRRRRRRRRRRRRRRRRRR

(R){20}
49: RRSRR RRRRRRRRRRRRRRRRRRR

(R){20}
50: RSRRR RRRRRRRRRRRRRRRRRRR

(R){20}
51: SRRRR RRRRRRRRRRRRRRRRRRR

ABG26513 ck: 3117 len: 265 ! Abg26513 Novel human diagnostic protein

(R,K){20}
199: EEEEE KKKKKKKKKKKKKKKKKKK

(K){20}

222: KKEEE KKKKKKKKKKKKKKKKK KKEE
(K)(20)
223: KEEKK KKKKKKKKKKKKKKKKK KKEE
(R,K)(20)
224: EEEKK KKKKKKKKKKKKKKKKK KEEEE
(R,K)(20)
225: EEEKK KKKKKKKKKKKKKKKKK EEEEE
(R,K)(20)
1
ABG26514 ck: 9106 len: 218 | Abg26514 Novel human diagnostic protein #26
(R,K)(20)
(R)(20)
182: EEEGG RRRRRRRRRRRRRRRRR RRRR
(R)(20)
183: ERRGR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
184: RRGRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
185: RGRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
186: GRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
187: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
188: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
189: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
190: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
191: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
192: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K)(20)
193: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R,K)(20)
1
ABG26515 ck: 7288 len: 389 | Abg26515 Novel human diagnostic protein #26
(R,K)(20)
(R,K)(20)
259: KKEES KKKKKKKKKKKKKKKKK KKKK
(K)(20)
260: KEEER KKKKKKKKKKKKKKKKK KKKR
(K)(20)
261: EESRR KKKKKKKKKKKKKKKKK KRRK
(K)(20)
262: ESRRK KKKKKKKKKKKKKKKKK KRRK
(R,K)(20)
263: SRKKK KKKKKKKKKKKKKKKKK KRRK
(R,K)(20)
264: RKKKK KKKKKKKKKKKKKKKKK KRRK
(R,K)(20)

265: KKKKK KKKKKKKKKKKKKKKKK KKKK
(R,K)(20)
1
ABG26516 ck: 2295 len: 91 | Abg26516 Novel human diagnostic protein
(R,K)(20)
(R,K)(20)
45: EEEEE RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
46: EEEER RRRRRRRRRRRRRRRRR KRRR
(R,K)(20)
47: EEEER RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
48: EEEER RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
49: EEEER RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
50: RRRRR RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
51: RRRRR RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
52: RRRRR RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
53: RRRRR RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
54: RRRRR RRRRRRRRRRRRRRRRR KRRK
(R,K)(20)
1
ABG26518 ck: 8431 len: 761 | Abg26518 Novel human diagnostic protein
(R,K)(20)
(R)(20)
725: EEEEE RRRRRRRRRRRRRRRRR RRRR
(R)(20)
726: EEEER RRRRRRRRRRRRRRRRR RRRR
(R)(20)
727: EEEER RRRRRRRRRRRRRRRRR RRRR
(R)(20)
728: EEEER RRRRRRRRRRRRRRRRR RRRR
(R)(20)
729: EEEER RRRRRRRRRRRRRRRRR RRRR
(R)(20)
730: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
731: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
732: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
733: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
734: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)
735: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R)(20)

736: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK
737: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK
738: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKKK
739: RRRR (R,K){20} RRRRRRRRRRRRRRRR KKK
740: RRRR (R,K){20} RRRRRRRRRRRRRRRR KN
741: RRRR (R,K){20} RRRRRRRRRRRRRRRR N

ABG26520 ck: 2487 len: 99 ! Abg26520 Novel human diagnostic protein #26

1

35: EEEE (R,K){20} RRRRRRRRRRRRRRRR RRRR
36: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR
37: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR
38: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR
39: EEEE (R){20} RRRRRRRRRRRRRRRR RRRR
40: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
41: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
42: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
43: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
44: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
45: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
46: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
47: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
48: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
49: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
50: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
51: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

52: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
53: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
54: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
55: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
56: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
57: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
58: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
59: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
60: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
61: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
62: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
63: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
64: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
65: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
66: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
67: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
68: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
69: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
70: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
71: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
72: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
73: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
74: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR
75: RRRR (R){20} RRRRRRRRRRRRRRRR RRRR

```
(R){20}
76: RRRR RRRRRRRRRRRRRRRRRR RRKI
(R){20}
77: RRRR RRRRRRRRRRRRRRRRRR RKI
(R){20}
78: RRRR RRRRRRRRRRRRRRRRRR KI
(R,K){20}
79: RRRR RRRRRRRRRRRRRRRRRR I

ABG26521 ck: 482 len: 367 ! Abg26521 Novel human diagnostic protein #26
(R,K){20}
317: EEEG KKKKKKKKKKKKKKKKK KKKR
(K){20}
318: EEEG KKKKKKKKKKKKKKKKK KKKR
(K){20}
319: EEGK KKKKKKKKKKKKKKKKK KRRR
(K){20}
320: EGKK KKKKKKKKKKKKKKKKK KRRK
(K){20}
321: GKKK KKKKKKKKKKKKKKKKK RRRK
(R,K){20}
322: KKKK KKKKKKKKKKKKKKKKK RRRK
(R,K){20}
323: KKKK KKKKKKKKKKKKKKKKK RKKK
(R,K){20}
324: KKKK KKKKKKKKKKKKKKKKK KKKK
(R,K){20}
325: KKKK KKKKKKKKKKKKKKKKK KKEQ
(R,K){20}
326: KKKK KKKKKKKKKKKKKKKKK KKEQ
(R,K){20}
327: KKKK KKKKKKKKKKKKKKKKK KEQOL
(R,K){20}
328: KKKK KKKKKKKKKKKKKKKKK EQOLE

ABG26522 ck: 2060 len: 152 ! Abg26522 Novel human diagnostic protein #26
(R,K){20}
46: EEEE RRRRRRRRRRRRRRRRR RKKKE
(R){20}
47: EEEER RRRRRRRRRRRRRRRRR RKKEG
(R){20}
48: EEEER RRRRRRRRRRRRRRRRR KKEGE
(R,K){20}
49: EEEER RRRRRRRRRRRRRRRRR KEGEE
(R,K){20}
50: EERRR RRRRRRRRRRRRRRRRRK EGEEG
(R,K){20}
96: KKEEG RRRRRRRRRRRRRRRRR RRRR
```

```
(R){20}
97: KEEGR RRRRRRRRRRRRRRRRR RRRR
(R){20}
98: EEGR RRRRRRRRRRRRRRRRR RRRR
(R){20}
99: EGRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
100: GRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
101: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
102: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
103: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
104: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
105: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
106: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
107: RRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
108: RRRR RRRRRRRRRRRRRRRRR RRRK
(R){20}
109: RRRR RRRRRRRRRRRRRRRRR RRRK
(R){20}
110: RRRR RRRRRRRRRRRRRRRRR RRRK
(R){20}
111: RRRR RRRRRRRRRRRRRRRRR RRRK
(R){20}
112: RRRR RRRRRRRRRRRRRRRRR RRRK
(R,K){20}
113: RRRR RRRRRRRRRRRRRRRRR RRRK

ABG26525 ck: 5433 len: 103 ! Abg26525 Novel human diagnostic protein
(R,K){20}
3: RD RRRKKKKKKKKKKKKKKKK EEEE

ABG26526 ck: 1887 len: 115 ! Abg26526 Novel human diagnostic protein
(R,K){20}
48: EEEE RRRRRRRRRRRRRRRRR KRRK
(R){20}
49: EEEER RRRRRRRRRRRRRRRRR KREK
(R,K){20}
50: EEEER RRRRRRRRRRRRRRRRR RRRK
(R,K){20}
```


51: EERRR RRRRRRRRRRRRRRRRKKR EKKKS
ABG26527 ck: 5586 len: 122 1 Abg26527 Novel human diagnostic protein #26
(R,K){20}
82: EEEEE KKKKKKKKKKKKKRRRRR GRRRM
(R,K){20}
1 ABG26528 ck: 2237 len: 215 1 Abg26528 Novel human diagnostic protein #26
(R,K){20}
87: EKEKE KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
88: KEKEK KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
89: EKEKK KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
90: KEKKR KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
91: EKKR KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
92: KKKR KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
93: KKKK KKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
94: KKKR KKKKKKKKKKKKKKKKKK KKKKK
(K){20}
95: KKKR KKKKKKKKKKKKKKKKKK KKKKK
(K){20}
96: KKKK KKKKKKKKKKKKKKKKKK EEEEE
(K){20}
167: GRRG RRRRRRRRRRRRRRRR RRRR
(R){20}
168: RRRG RRRRRRRRRRRRRRRR RRRR
(R){20}
169: RRRG RRRRRRRRRRRRRRRR RRRR
(R){20}
170: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
171: GRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
172: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
173: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
174: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
175: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
176: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}

177: RRRR RRRRRRRRRRRRRRRR MRKR
(R){20}
ABG26530 ck: 5729 len: 404 1 Abg26530 Novel human diagnostic protein
(R,K){20}
366: EEEEE RRRRRRRRRRRRRRRR RRRR
(R){20}
367: EEEER RRRRRRRRRRRRRRRR RRRR
(R){20}
368: EEEER RRRRRRRRRRRRRRRR RRRR
(R){20}
369: EERRR RRRRRRRRRRRRRRRR RRRR
(R){20}
370: ERRR RRRRRRRRRRRRRRRR RRRR
(R){20}
371: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
372: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
373: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
374: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
375: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
376: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
377: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
378: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
379: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
1 ABG26531 ck: 7434 len: 126 1 Abg26531 Novel human diagnostic protein
(R,K){20}
48: EEKDE RRRRRRRRRRRRRRRR RRRR
(R){20}
49: EKDER RRRRRRRRRRRRRRRR RRRR
(R){20}
50: KDEER RRRRRRRRRRRRRRRR RRRR
(R){20}
51: DEERR RRRRRRRRRRRRRRRR RRRR
(R){20}
52: ERRR RRRRRRRRRRRRRRRR RRRR
(R){20}
53: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
54: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}

47: KKKK KKKKKKKKKKKKKKK KKKK
(K){20}
48: KKKK KKKKKKKKKKKKKKK KKKK
(K){20}
49: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
50: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
51: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
52: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
53: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
54: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
55: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
ABG26535 ck: 2214 len: 472 1 Abg26535 Novel human diagnostic protein #26

1
415: EEEE RRRRRRRRRRRRRRR RRRR
(R){20}
416: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
417: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
418: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
419: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
420: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
421: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
422: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
423: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
424: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
425: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
426: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
427: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}
428: EEEER RRRRRRRRRRRRRRR RRRR
(R){20}

429: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
430: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
431: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
432: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
433: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
434: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
435: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
436: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
437: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}
438: RRRR RRRRRRRRRRRRRRR RRRR
(R){20}

1
ABG26537 ck: 2429 len: 573 1 Abg26537 Novel human diagnostic protein
(R,K){20}
405: GAGE RRRRRRRRRRRRRRR RREE
(R){20}
406: AEOER RRRRRRRRRRRRRRR RREE
(R){20}
407: EOER RRRRRRRRRRRRRRR RREE
(R){20}
479: KKEE KKKKKKKKKKKKKKK KKKK
(R,K){20}
480: KKEE KKKKKKKKKKKKKKK KKKK
(R,K){20}
481: EEEK KKKKKKKKKKKKKKK KKKK
(R,K){20}
482: EEEK KKKKKKKKKKKKKKK KKKK
(R){20}
483: EKKK KKKKKKKKKKKKKKK KKKK
(R){20}
484: KKKK KKKKKKKKKKKKKKK KKKK
(R){20}
485: KKKK KKKKKKKKKKKKKKK KKKK
(R){20}
486: KKKK KKKKKKKKKKKKKKK KKKK
(R){20}

1
ABG26538 ck: 5732 len: 98 1 Abg26538 Novel human diagnostic protein
(R,K){20}
60: KEKE KRRRRRRRRRRRRRR KKKK
(R,K){20}

[illegible][illegible]

34:	RRRRR	(R) [20]	RRRRR
35:	RRRRR	(R) [20]	RRRRR
36:	RRRRR	(R) [20]	RRRRR
37:	RRRRR	(R) [20]	RRRRR
38:	RRRRR	(R) [20]	RRRRR
39:	RRRRR	(R) [20]	RRRRR
40:	RRRRR	(R) [20]	RRRRR
41:	RRRRR	(R) [20]	RRRRR
42:	RRRRR	(R) [20]	RRRRR
43:	RRRRR	(R) [20]	RRRRR
44:	RRRRR	(R) [20]	RRRRR
45:	RRRRR	(R) [20]	RRRRR
46:	RRRRR	(R) [20]	RRRRR
47:	RRRRR	(R) [20]	RRRRR
48:	RRRRR	(R) [20]	RRRRR
49:	RRRRR	(R) [20]	RRRRR
50:	RRRRR	(R) [20]	RRRRR

```

ABG26542  ck: 4605  len: 182  ! Abg26542 Novel human diagnostic protein
      (R,K)(20)
146: EGEDE RRRRRRRRRRRRRRRRRRRR RRRGR
      (R)(20)
147: GEEER RRRRRRRRRRRRRRRRRRRR RRRGR
      (R)(20)
148: ZEERR RRRRRRRRRRRRRRRRRRRR RGRRG
      (R)(20)
149: ZEERR RRRRRRRRRRRRRRRRRRRR GRGRG
      (R)(20)
ABG26543  ck: 2898  len: 405  ! Abg26543 Novel human diagnostic protein
      (R,K)(20)
264: RRRRW RRRRRRRRRRRRRRRRRRRR RRRRR
      (R,K)(20)

```

[illegible][illegible]

18: KKKKK (R,K){20}
19: KKKKK (R,K){20}
20: KKKKK (R,K){20}
21: KKKKK (R,K){20}

ABG26717 ck: 4415 len: 78 ! Abg26717 Novel human diagnostic protein #26

37: QMLSV (R,K){20}
38: MLSVK (R,K){20}
39: LSVKK (R,K){20}
40: SVKKK (R,K){20}
41: VKKKK (R,K){20}
42: KKKKK (R,K){20}
43: KKKKK (K){20}
44: KKKKK (K){20}
45: KKKKK (K){20}
46: KKKKK (K){20}
47: KKKKK (K){20}
48: KKKKK (K){20}
49: KKKKK (K){20}
50: KKKKK (K){20}
51: KKKKK (K){20}
52: KKKKK (K){20}
53: KKKKK (K){20}
54: KKKKK (K){20}

ABG26718 ck: 9531 len: 141 ! Abg26718 Novel human diagnostic protein #26

1 (R,K){20}

80: EVARP (R,K){20}
81: VARP (R,K){20}
82: ARPRK (K){20}
83: RPRKK (K){20}
84: PRKKK (K){20}
85: RKKKK (K){20}
86: KKKKK (K){20}
87: KKKKK (K){20}
88: KKKKK (K){20}
89: KKKKK (K){20}

ABG26719 ck: 72 len: 83 ! Abg26719 Novel human diagnostic protein

42: ETPSE (R,K){20}
43: TPSEK (K){20}
44: PSEKK (K){20}
45: SEKKK (K){20}
46: EKKKK (K){20}
47: KKKKK (K){20}
48: KKKKK (K){20}
49: KKKKK (K){20}
50: KKKKK (K){20}
51: KKKKK (K){20}
52: KKKKK (K){20}
53: KKKKK (K){20}

1
ABG26720 ck: 3781 len: 57 ! Abg26720 Novel human diagnostic protein #26

54: KKKK KKKKKKKKKKKKKKKKKKKKKK NPIFF

(R,K){20}

24: QMKST KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: MKSIK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KSIIK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: SIKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: IKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKKKKK KEHFL

(K){20}

33: KKKKK KKKKKKKKKKKKKKKKKKKKKK FHEFL

(K){20}

ABG26721 ck: 287 len: 95 ! Abg26721 Novel human diagnostic protein #26

1

(R,K){20}

31: EKEKE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: KEKEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: EKEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: KEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

42: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKKKKK NKKKK

(K){20}

ABG26722 ck: 4831 len: 127 ! Abg26722 Novel human diagnostic protein

1

(R,K){20}

44: KKKKE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: KKEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KEEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: EKKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKKKKK KEKKK

(K){20}

87: EKEEE KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

88: KEEEK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

89: EEEKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

90: EEKKK KKKKKKKKKKKKKKKKKKKKKK KKKKK

91: EKKKK (K){20} KKKKK
92: KKKKK (K){20} KKKKK
93: KKKKK (K){20} KKKKK
94: KKKKK (K){20} KKKKK
95: KKKKK (K){20} KKKKK
96: KKKKK (K){20} KKKKK
97: KKKKK (K){20} KKKKK
98: KKKKK (K){20} KKKKK
99: KKKKK (K){20} KKKKK
100: KKKKK (K){20} KKKKK
101: KKKKK (K){20} KKKKK
102: KKKKK (K){20} KKKKK
103: KKKKK (K){20} KKKKK
104: KKKKK (K){20} KKKKK
105: KKKKK (K){20} KKKKK
106: KKKKK (K){20} KKKKK
107: KKKKK (K){20} KKKKK
108: KKKKK (K){20} KKKKK
ABG26723 ck: 7054 len: 98 ! Abg26723 Novel human diagnostic protein #26
(R,K){120}
42: EKEKE (K){20} KKKKK
43: KEKEK (K){20} KKKKK
44: EKEKK (K){20} KKKKK
45: KEKKK (K){20} KKKKK
46: EKKKK (K){20} KKKKK

47: KKKKK (K){20} KKKKK
48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
50: KKKKK (K){20} KKKKK
51: KKKKK (K){20} KKKKK
52: KKKKK (K){20} KKKKK
53: KKKKK (K){20} KKKKK
54: KKKKK (K){20} KKKKK
55: KKKKK (K){20} KKKKK
56: KKKKK (K){20} KKKKK
57: KKKKK (K){20} KKKKK
58: KKKKK (K){20} KKKKK
59: KKKKK (K){20} KKKKK
60: KKKKK (K){20} KKKKK
61: KKKKK (K){20} KKKKK
62: KKKKK (K){20} KKKKK
63: KKKKK (K){20} KKKKK
64: KKKKK (K){20} KKKKK
65: KKKKK (K){20} KKKKK
66: KKKKK (K){20} KKKKK
67: KKKKK (K){20} KKKKK
68: KKKKK (K){20} KKKKK
69: KKKKK (K){20} KKKKK
70: KKKKK (K){20} KKKKK

1

ABG26724 ck: 9897 len: 295 1 Abg26724 Novel human diagnostic protein #26

(R,K){20}

259: KEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

260: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

261: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

262: EEEER KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

263: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

264: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

265: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

266: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

267: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

268: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

269: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

270: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

271: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

272: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

273: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

274: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

275: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

276: KKKKK KKKKKKKKKKKKKKKKKKK K

1

ABG26725 ck: 909 len: 131 1 Abg26725 Novel human diagnostic protein #26

(R,K){20}

70: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

71: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

72: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

73: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

1

ABG26726 ck: 9300 len: 677 1 Abg26726 Novel human diagnostic protein

(R,K){20}

240: EKEKE KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

241: KEKER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

242: EKEER KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

243: KEERK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

244: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

245: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

246: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

247: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R,K){20}

248: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

249: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

250: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

251: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

252: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

```

253: RKKKK (K){20}
254: KKKKK (K){20}
255: KKKKK (K){20}
256: KKKKK (K){20}
257: KKKKK (K){20}
258: KKKKK (K){20}
259: KKKKK (K){20}
260: KKKKK (K){20}
261: KKKKK (R,K){20}
262: KKKKK (R,K){20}
263: KKKKK (R,K){20}
264: KKKKK (R,K){20}
265: KKKKK (R,K){20}

ABG26727 ck: 813 len: 329 1 Abg26727 Novel human diagnostic protein #26
1
215: EKEKE (R,K){20}
216: KEKEK (K){20}
217: EKEKK (K){20}
218: KEKKK (K){20}
219: EKKKK (K){20}
220: KKKKK (K){20}
221: KKKKK (K){20}
222: KKKKK (K){20}
223: KKKKK (K){20}
224: KKKKK (K){20}

```

```

225: KKKKK (K){20}
226: KKKKK (K){20}
227: KKKKK (K){20}
228: KKKKK (K){20}
229: KKKKK (K){20}
230: KKKKK (K){20}

ABG26879 ck: 2669 len: 800 1 Abg26879 Novel human diagnostic protein
1
559: EKEKE (R,K){20}
560: KEKEK (R,K){20}
561: EKEKK (R,K){20}
562: KEKKK (K){20}
563: EKKKK (K){20}
564: KKKKK (K){20}

ABG28885 ck: 8668 len: 137 1 Abg28885 Novel human diagnostic protein
1
35: EEEEG (R,K){20}
36: EEEGR (R,K){20}
37: EEEGR (R){20}
38: EGRRR (R){20}
39: GRRRR (R){20}
40: RRRRR (R,K){20}
41: RRRRR (R,K){20}

AAG65985 ck: 8085 len: 154 1 Aag65985 B726P splice variant sequence.
1
114: TQLRQ (K){20}

```

115: QLRQK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
116: LRQKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
117: RQKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
118: OKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
119: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
120: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
121: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
122: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
123: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
124: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
125: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
126: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
127: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
128: KKKKK KKKKKKKKKKKKKKKKKK KKKKK
(K,120)
ABb27893 Ck: 5383 len: 86 1 Abb27893 Human peptide #544 encoded by brea
(R,K,120)
15: RRRRG RRRRRKKKKKKKKKKKKK KRRRR
(R,K,120)
16: RRRGR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
17: RRRGR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
18: RGRGR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
19: GRRRR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
20: RRRRR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
21: RRRRR RRRKKKKKKKKKKKKKKK KRRRR
(R,K,120)
22: RRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
23: RRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)

1
24: RRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
25: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
26: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
27: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
28: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
29: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
30: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
31: RRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
32: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
33: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
34: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
35: KRRKK KKKKKKKKKKKKKKKKKK RRRRR
(R,K,120)
36: KRRKK KRRRRRRRRRRRRRRRRR RRRRR
(R,K,120)
37: RRRKK RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
38: KRRKK RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
39: KRRKK RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
40: KRRKK RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
41: KRRKK RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
42: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
43: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
44: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
45: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,120)
ABb28750 Ck: 1334 len: 86 1 Abb28750 Peptide #1401 encoded by brea
(R,K,120)
57: EEEEG RRRKKKKKKKKKKKKKKK KKKKK
(R,K,120)

1

```
45: KERRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
46: ERRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
47: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
48: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
49: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
50: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
51: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
52: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
53: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}
54: FRRRR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (R,K){20}

ABB29928 ck: 3607 len: 88 ! Abb29928 Peptide #2579 encoded by breast
      (R,K){20}
      (K){20}
39: EKKRE RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
40: RKRER RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
41: KRERK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
42: REKKR RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
43: EKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
44: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
45: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
46: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
47: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
48: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
49: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
50: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
51: KKKKK RRRRRRRRRRRRRRRRRRRR RRRKK (R,K){20}
      (K){20}
```

52: KKKK (K){20} KKKK
53: KKKK (K){20} KKKK
54: KKKK (K){20} KKKK
55: KKKK (K){20} KKKK
56: KKKK (K){20} KKKK
57: KKKK (K){20} KKKK
58: KKKK (K){20} KKKK
59: KKKK (K){20} KKKK
60: KKKK (K){20} KKKK
61: KKKK (K){20} KKKK
62: KKKK (K){20} KKKK
63: KKKK (K){20} KKKK
64: KKKK (K){20} KKKK
65: KKKK (K){20} KKKK
66: KKKK (K){20} KKKK
67: KKKK (K){20} KKKK
68: KKKK (K){20} K
69: KKKK (K){20} KKKK
AB030512 ck: 3937 len: 85 ! Ab030512 Peptide #3163 encoded by breast c

6: KKKK (K){20} KKKK
7: KKKK (K){20} KKKK
8: KKKK (K){20} KKKK
9: KKKK (K){20} KKKK
10: KKKK (K){20} KKKK
11: KKKK (K){20} KKKK
AB032308 ck: 1560 len: 88 ! Ab032308 Peptide #4959 encoded by breas

43: RRRG (R){20} RRRR
44: ERGR (R){20} RRRR
45: RGRG (R){20} RRRR
46: RGRG (R){20} RRRR
47: GRGR (R){20} RRRR
48: RRRR (R){20} RRRR
49: RRRR (R){20} RRRR
50: RRRR (R){20} RRRR
51: RRRR (R){20} RRRR
52: RRRR (R){20} RRRR
53: RRRR (R){20} RRRR
54: RRRR (R){20} RRRR
55: RRRR (R){20} RRRR
56: RRRR (R){20} RRRR
57: RRRR (R){20} RRRR
58: RRRR (R){20} RRRR
59: RRRR (R){20} RRRR

60: RRRR (R){20} RRRRRRRRRRRRRRRRRRR RRRTN
61: RRRR (R){20} RRRRRRRRRRRRRRRRRRR RNTNN
62: RRRR (R){20} RRRRRRRRRRRRRRRRRRR NTNNE
AB33064 ck: 5383 len: 86 i Ab33064 Peptide #570 encoded by human foet
15: RRRG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
16: RRRG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
17: RRRG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
18: RRRG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
19: RRRG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
20: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
21: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
22: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
23: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
24: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
25: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
26: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
27: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
28: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
29: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
30: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
31: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
32: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
33: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
34: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR

1
35: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
36: RRRR (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
37: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
38: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
39: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
40: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
41: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
42: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
43: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
44: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
45: RRRR (R){20} RRRRRRRRRRRRRRRRRRR KRRR
AB33937 ck: 1334 len: 86 i Ab33937 Peptide #1443 encoded by human
57: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
58: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
59: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
60: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
61: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
62: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
63: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
64: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
65: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
66: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
67: EEEG (R,K){20} RRRRRRRRRRRRRRRRRRR KRRR
AB34024 ck: 9082 len: 167 i Ab34024 Peptide #1530 encoded by human

1

(R,K){20}
(R,K){20}
33: EEGRG RRRRRRRRRRRRRRRRRR RRGCG
(R,K){20}
34: EGRGR RRRRRRRRRRRRRRRRRR RGGGR
(R,K){20}
35: GRGRR RRRRRRRRRRRRRRRRRR GGGRR

1

ABB34533 ck: 2276 len: 89 1 Abb34533 Peptide #2039 encoded by human fo
(R,K){20}
(R,K){20}
23: EEEEE KKKKKKKKKKKKKKKKK EEEK
(K){20}
46: KKEEE KKKKKKKKKKKKKKKKK KRKEE
(K){20}
47: KEEER KKKKKKKKKKKKKKKKK KKEEE
(R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKK KEEEE
(R,K){20}
49: EEEKK KKKKKKKKKKKKKKKKK EEEEE

1

ABB34819 ck: 1939 len: 130 1 Abb34819 Peptide #2325 encoded by human fo
(R,K){20}
(R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
43: GRKER RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
44: KERRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
45: KERRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
46: EERRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
47: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
48: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
49: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
50: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
51: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
52: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
53: RRRRR RRRRRRRRRRRRRRRRRR RRRRR
(R,K){20}
54: RRRRR RRRRRRRRRRRRRRRRRR RRRRR

1

ABB5110 ck: 3607 len: 88 1 Abb5110 Peptide #2616 encoded by human
(R,K){20}
(K){20}
39: ERRRE KKKKKKKKKKKKKKKKK KRRKK
(K){20}
40: KRREK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
41: KREKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
42: REKKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
43: EKRRK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
44: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
45: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
46: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
47: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
48: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
49: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
50: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
51: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
52: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
53: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
54: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
55: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
56: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
57: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
58: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
59: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
60: KRRKK KKKKKKKKKKKKKKKKK KRRKK
(K){20}
61: KRRKK KKKKKKKKKKKKKKKKK KRRKK

62: KKKKK (K)[20] KKKKK
63: KKKKK (K)[20] KKKKK
64: KKKKK (K)[20] KKKKK
65: KKKKK (K)[20] KKKKK
66: KKKKK (K)[20] KKKKK
67: KKKKK (K)[20] KKKKK
68: KKKKK (K)[20] KKKKK
69: KKKKK (K)[20] KKKKK
ABB35676 ck: 3937 len: 85 ! Abb35676 Peptide #3182 encoded by human foa
(R,K)[20]
1: KKKKK (K)[20] KKKKK
2: KKKKK (K)[20] KKKKK
3: KKKKK (K)[20] KKKKK
4: KKKKK (K)[20] KKKKK
5: KKKKK (K)[20] KKKKK
6: KKKKK (K)[20] KKKKK
7: KKKKK (K)[20] KKKKK
8: KKKKK (K)[20] KKKKK
9: KKKKK (K)[20] KKKKK
10: KKKKK (K)[20] KKKKK
11: KKKKK (K)[20] KKKKK
ABB36406 ck: 2686 len: 71 ! Abb36406 Peptide #3912 encoded by human foa
(R,K)[20]
20: KKKKK (K)[20] KKKKK
21: KKKKK (K)[20] KKKKK
22: KKKKK (K)[20] KKKKK

23: KKKKK (K)[20] KKKKK
24: KKKKK (K)[20] KKKKK
25: KKKKK (R,K)[20] KKKKK
26: KKKKK (R,K)[20] KKKKK
27: KKKKK (R,K)[20] KKKKK
28: KKKKK (R,K)[20] KKKKK
29: KKKKK (R,K)[20] KKKKK
30: KKKKK (R,K)[20] KKKKK
31: KKKKK (R,K)[20] KKKKK
32: KKKKK (R,K)[20] KKKKK
33: KKKKK (R,K)[20] KKKKK
34: KKKKK (R,K)[20] KKKKK
35: KKKKK (R,K)[20] KKKKK
36: KKKKK (R,K)[20] KKKKK
37: KKKKK (R,K)[20] KKKKK
38: KKKKK (R,K)[20] KKKKK
39: KKKKK (R,K)[20] KKKKK
40: KKKKK (R,K)[20] KKKKK
41: KKKKK (R,K)[20] KKKKK
42: KKKKK (R,K)[20] KKKKK
43: KKKKK (R,K)[20] KKKKK
44: KKKKK (R,K)[20] KKKKK
45: KKKKK (R,K)[20] KKKKK
46: KKKKK (R,K)[20] KKKKK

(K){20}
47: KKKR KKKKKKKKKKKKKKK KSAH
(K){20}
48: KKKR KKKKKKKKKKKKKKK KSAH
(K){20}
49: KKKR KKKKKKKKKKKKKKK SAH
AB37567 ck: 1560 len: 88 1 Ab37567 Peptide #5073 encoded by human fo
(R,K){20}
43: RRRG RRRRRRRRRRRRRRRR RRRR
(R){20}
44: ERGR RRRRRRRRRRRRRRRR RRRR
(R){20}
45: RGRR RRRRRRRRRRRRRRRR RRRR
(R){20}
46: RGRR RRRRRRRRRRRRRRRR RRRR
(R){20}
47: GRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
48: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
49: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
50: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
51: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
52: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
53: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
54: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
55: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
56: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
57: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
58: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
59: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
60: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
61: RRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
62: RRRR RRRRRRRRRRRRRRRR RRRR

1
AB37780 ck: 2324 len: 36 1 Ab37780 Peptide #5286 encoded by human
(R,K){20}
16: KRRR KRRRRRRRRRRRRRRR R
(R,K){20}
17: ERRT KRRRRRRRRRRRRRRR
AB40272 ck: 8343 len: 66 1 Ab40272 Peptide #7778 encoded by human
(R,K){20}
6: EERE KRRRRRRRRRRRRRRR RRRR
(R,K){20}
7: TERE KRRRRRRRRRRRRRRR RRRR
(R,K){20}
8: EERE KRRRRRRRRRRRRRRR RRRR
(R,K){20}
9: REKK KRRRRRRRRRRRRRRR RRRR
AB42642 ck: 2394 len: 57 1 Ab42642 Peptide #10148 encoded by huma
(R,K){20}
20: EEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
21: EEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
22: EEEG RRRRRRRRRRRRRRRR RRRR
(R){20}
23: EGRR RRRRRRRRRRRRRRRR RRRR
(R){20}
24: GRRR RRRRRRRRRRRRRRRR RRRR
(R){20}
25: RRRR RRRRRRRRRRRRRRRR RRRR
AB43181 ck: 4228 len: 24 1 Ab43181 Peptide #10687 encoded by huma
(R,K){20}
1: RRRR RRRRRRRRRRRRRRRR RRRR
(R,K){20}
2: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}
3: R RRRRRRRRRRRRRRRR RRRR
(R,K){20}
4: RRR RRRRRRRRRRRRRRRR RRRR
AB44317 ck: 4895 len: 51 1 Ab44317 Peptide #11823 encoded by huma
(R,K){20}
18: LKPM RRRRRRRRRRRRRRRR RRRR
(R,K){20}
19: FKPM RRRRRRRRRRRRRRRR RRRR

(R,K){20}
20: KPWKK RRRKKRRRRKKRRKKK LTTT

1
ABBI7165 ck: 8887 len: 42 1 Abbi7165 Human nervous system related polyA
(R,K){20}
(K){20}
21: FTTTE KKKKKKKKKKKKKKKK KX
(K){20}
22: FTTEK KKKKKKKKKKKKKKKK X

1
ABBI8534 ck: 5383 len: 86 1 Abbi8534 Protein #533 encoded by probe for
(R,K){20}
(R,K){20}
15: RRRRG RRRRRKKKKKKKKKKKK KRRR
(R,K){20}
16: RRRGR RRRKKKKKKKKKKKKKK KRRR
(R,K){20}
17: RRGRR RRRKKKKKKKKKKKKKK RRRR
(R,K){20}
18: RGRRR RRRKKKKKKKKKKKKKK RRRR
(R,K){20}
19: GRRRR RKKKKKKKKKKKKKKKK RRRR
(R,K){20}
20: RRRRR KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
21: RRRRK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
22: RRRKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
23: RRRKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
24: RKKKK KRRKKKKKKKKKKKKKK RRRR
(R,K){20}
25: KKKKK KRRKKKKKKKKKKKKKK RRRR
(R,K){20}
26: KKKKK RKKKKKKKKKKKKKKKK RRRR
(R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
29: KRRKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
30: KRRKK KKKKKKKKKKKKKKKKK RRRR
(R,K){20}
31: RKKKK KRRKKKKKKKKKKKKKK RRRR
(R,K){20}
32: KKKKK RKKKKKKKKKKKKKKKK RRRR
(R,K){20}

1
33: KKKKK KKKKKRRRRRRRRRRRR RRRR
(R,K){20}
34: KKKKK KKKRRRRRRRRRRRRRR RRRR
(R,K){20}
35: KKKKK KRRRRRRRRRRRRRRRR RRRR
(R,K){20}
36: KKKKK KRRRRRRRRRRRRRRRR RRRR
(R){20}
37: RKKKK RRRRRRRRRRRRRRRRR RRRR
(R){20}
38: KKKKK RRRRRRRRRRRRRRRRR RRRR
(R){20}
39: KKKRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
40: KKKRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
41: KRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
42: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
43: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
44: RRRRR RRRRRRRRRRRRRRRRR RRRR
(R){20}
45: RRRRR RRRRRRRRRRRRRRRRR RRRR

1
ABBI9373 ck: 1334 len: 86 1 Abbi9373 Protein #1372 encoded by probe
(R,K){20}
(R,K){20}
57: EEEEG RRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
58: EEEGR RRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
59: EEEGR RRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
60: EGGRR KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
61: GRRRK KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
62: RRRKK KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
63: RRRKK KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
64: RKKKK RRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
65: KKKRR KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
66: KKKKK RRRKKKKKKKKKKKKKK KKKKK
(K){20}

67: KKKR KKKKKKKKKKKKKKK

ABb19467 ck: 9082 len: 167 ! Abb19467 Protein #1466 encoded by probe for

(R,K){20}
(R,K){20}

33: EGGG RRRRRRRRRRRRRRRR RGGG

(R,K){20}
34: EGGG RRRRRRRRRRRRRRRR RGGG

(R,K){20}
35: GGGG RRRRRRRRRRRRRRRR GGGG

ABb19943 ck: 2276 len: 89 ! Abb19943 Protein #1942 encoded by probe for

(R,K){20}
(R,K){20}

23: EEEE KKKKKKKKKKKKKKKK EEEK

(K){20}
46: KKEE KKKKKKKKKKKKKKKK KKEE

(K){20}
47: KKEE KKKKKKKKKKKKKKKK KKEE

(R,K){20}
48: EEEK KKKKKKKKKKKKKKKK EEEE

(R,K){20}
49: EEEK KKKKKKKKKKKKKKKK EEEE

ABb20238 ck: 1939 len: 130 ! Abb20238 Protein #2237 encoded by probe for

(R,K){20}
(R,K){20}

42: EGGE RRRRRRRRRRRRRRRR RRRR

(R,K){20}
43: GKEE RRRRRRRRRRRRRRRR RRRR

(R,K){20}
44: KEER RRRRRRRRRRRRRRRR RRRR

(R,K){20}
45: KEER RRRRRRRRRRRRRRRR RRRR

(R,K){20}
46: EERR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
47: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
48: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
49: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
50: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
51: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
52: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}
53: RRRR RRRRRRRRRRRRRRRR RRRR

(R,K){20}

54: RRRR RRRRRRRRRRRRRRRR EEEE

ABb20531 ck: 3607 len: 88 ! Abb20531 Protein #2530 encoded by probe

(R,K){20}
(K){20}

39: ERRE KKKKKKKKKKKKKKKK KKKK

(K){20}
40: KKEE KKKKKKKKKKKKKKKK KKKK

(K){20}
41: KKEE KKKKKKKKKKKKKKKK KKKK

(K){20}
42: REKK KKKKKKKKKKKKKKKK KKKK

(K){20}
43: EKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
44: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
45: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
46: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
47: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
48: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
49: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
50: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
51: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
52: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
53: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
54: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
55: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
56: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
57: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
58: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
59: KKKK KKKKKKKKKKKKKKKK KKKK

(K){20}
60: KKKK KKKKKKKKKKKKKKKK KKKK

61: KKKK (K){20} KKKK
 62: KKKK (K){20} KKKK
 63: KKKK (K){20} KKKK
 64: KKKK (K){20} KKKK
 65: KKKK (K){20} KKKK
 66: KKKK (K){20} KKKK
 67: KKKK (K){20} KKKK
 68: KKKK (K){20} KKKK
 69: KKKK (K){20} KKKK

ABB21105 ck: 3937 len: 85 1 Abb21105 Protein #3104 encoded by probe for

1
 1: (R,K){20} KKKK
 2: (K){20} KKKK
 3: (K){20} KKKK
 4: (K){20} KKKK
 5: (K){20} KKKK
 6: (K){20} KKKK
 7: (K){20} KKKK
 8: (K){20} KKKK
 9: (K){20} KKKK
 10: (K){20} KKKK
 11: (K){20} KKKK
 ABB21763 ck: 2686 len: 71 1 Abb21763 Protein #3762 encoded by probe for
 (R,K){20} KKKK
 20: (K){20} KKKK

21: KKKK (K){20} KKKK
 22: KKKK (K){20} KKKK
 23: KKKK (K){20} KKKK
 24: KKKK (K){20} KKKK
 25: KKKK (R,K){20} KKKK
 26: KKKK (R,K){20} KKKK
 27: KKKK (R,K){20} KKKK
 28: KKKK (R,K){20} KKKK
 29: KKKK (R,K){20} KKKK
 30: KKKK (R,K){20} KKKK
 31: KKKK (R,K){20} KKKK
 32: KKKK (R,K){20} KKKK
 33: KKKK (R,K){20} KKKK
 34: KKKK (R,K){20} KKKK
 35: KKKK (R,K){20} KKKK
 36: KKKK (R,K){20} KKKK
 37: KKKK (R,K){20} KKKK
 38: KKKK (R,K){20} KKKK
 39: KKKK (R,K){20} KKKK
 40: KKKK (R,K){20} KKKK
 41: KKKK (R,K){20} KKKK
 42: KKKK (R,K){20} KKKK
 43: KKKK (R,K){20} KKKK
 44: KKKK (R,K){20} KKKK
 45: KKKK (R,K){20} KKKK

46: KKKK (R,K){20} KKKSA
47: KKKK (K){20} KKSAA
48: KKKK (K){20} KSAH
49: KKKK (K){20} SAH
AB22862 ck: 1560 len: 88 ! Ab22862 Protein #4861 encoded by probe for
(R,K){20}
43: EERG (R){20} RRRR
44: EERG (R){20} RRRR
45: EERG (R){20} RRRR
46: EERG (R){20} RRRR
47: EERG (R){20} RRRR
48: EERG (R){20} RRRR
49: EERG (R){20} RRRR
50: EERG (R){20} RRRR
51: EERG (R){20} RRRR
52: EERG (R){20} RRRR
53: EERG (R){20} RRRR
54: EERG (R){20} RRRR
55: EERG (R){20} RRRR
56: EERG (R){20} RRRR
57: EERG (R){20} RRRR
58: EERG (R){20} RRRR
59: EERG (R){20} RRRR
60: EERG (R){20} RRRR
61: EERG (R){20} RRRR

62: RRRR (R){20} NTNN
AB23064 ck: 2324 len: 36 ! Ab23064 Protein #5063 encoded by probe
(R,K){20}
16: KERT (R,K){20} R
17: ERTK (R,K){20} RRRR
AB24685 ck: 8343 len: 66 ! Ab24685 Protein #6684 encoded by probe
(R,K){20}
6: ETER (R,K){20} KKKK
7: TERK (R,K){20} KKKK
8: EREK (R,K){20} KKKK
9: REKK (R,K){20} KKKK
AB25988 ck: 2394 len: 57 ! Ab25988 Protein #7987 encoded by probe
(R,K){20}
20: EEEG (R,K){20} RRRR
21: EEEG (R){20} RRRR
22: EEEG (R){20} RRRR
23: EEEG (R){20} RRRR
24: EEEG (R){20} RRRR
25: EEEG (R){20} RRRR
AB27176 ck: 4895 len: 51 ! Ab27176 Protein #9175 encoded by probe
(R,K){20}
18: LEPK (R,K){20} KKKK
19: FKPM (R,K){20} KKKK
20: KPMK (R,K){20} KKKK
AB10296 ck: 3983 len: 292 ! Ab10296 Human cDNA SEQ ID NO: 604. 1/2
(R,K){20}
273: QVAP (R,K){20} KKKK

```
1 ABB10485   ck: 7611   len: 315   ! Abb10485 Human cDNA SEQ ID NO: 793. 1/2002
      (R,K){20}
273: QVEAP  RRRRRRRRRRRRRRRRRRRRRR KGGRS
      (K){20}
274: VFAPR  KKKKKKKKKKKKKKKKKKKKK GGRSR

1 AAU21948   ck: 444   len: 66   ! AAU21948 Human cardiovascular system antigen
      (R,K){20}
35: SMTPS  KKKKKKKKKKKKKKKKKKKKK KXGKK
      (K){20}
36: MTFBK  KKKKKKKKKKKKKKKKKKKKK XGKKK

1 AAU22148   ck: 4665  len: 34   ! AAU22148 Human cardiovascular system antigen
      (R,K){20}
10: PELIL  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
11: ELLLK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
12: LLLKK  KKKKKKKKKKKKKKKKKKKKK KKK
      (K){20}
13: LLKKK  KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
14: LKKKK  KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
15: KKKKK  KKKKKKKKKKKKKKKKKKKKK

1 AAU22186   ck: 269   len: 76   ! AAU22186 Human cardiovascular system antigen
      (R,K){20}
37: TPSRA  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
38: PSRAK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
39: SRAKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
40: RAKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
41: AKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU22374   ck: 8278  len: 53   ! AAU22374 Human cardiovascular system antigen
      (R,K){20}

1
30: NCGIL  KKKKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
31: CGILK  KKKKKKKKKKKKKKKKKKKKK KKK
      (K){20}
32: GILKK  KKKKKKKKKKKKKKKKKKKKK KK
      (K){20}
33: ILKKK  KKKKKKKKKKKKKKKKKKKKK K
      (K){20}
34: LKKKK  KKKKKKKKKKKKKKKKKKKKK

1 AAU23799   ck: 6158  len: 272   ! AAU23799 Novel human enzyme polypeptide
      (R,K){20}
238: SFANA  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
239: PANAK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
240: ANAKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
241: NAKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
242: AKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU27944   ck: 1121  len: 69   ! AAU27944 Human config polypeptide sequen
      (R,K){20}
40: VPPLT  RRRRRRRRRRRRRRRRRRRRRR KKKKK
      (R,K){20}
41: PPLTR  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
42: PLTRK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
43: LTRKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
44: TRKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
45: RKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
46: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK
      (R,K){20}
47: KKKKK  KKKKKKKKKKKKKKKKKKKKK KKKKK

1 AAU31467   ck: 4264  len: 657   ! AAU31467 Novel human secreted protein #1
      (R,K){20}
19: RRRRP  RRRRRRRRRRRRRRRRRRRRRR RRRRL
      (R){20}
20: RRRRP  RRRRRRRRRRRRRRRRRRRRRR RRRRL
      (R){20}
```

21: RPRRR RRRRRRRRRRRRRRRRRR RLGL
(R){20}
22: RPRRR RRRRRRRRRRRRRRRRRR RLGLE
(R){20}
23: RRRRR RRRRRRRRRRRRRRRRRR LGLR
AAU3348 ck: 8085 len: 154 1 Aau3348 Human breast cancer protein encode
(R,K){20}
114: TQLRQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
115: QLROK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
116: LROKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
117: RQKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
118: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
AAE09664 ck: 1663 len: 87 1 Aae09664 Human pancreatic related protein H
(R,K){20}
36: KMSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: WSSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: SSXKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: SXKKK KKKKKKKKKKKKKKKKKKK KKKKK

40: XKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
63: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

1 AAM95365 ck: 5626 len: 139 ! Aam95365 Human reproductive system related
(R,K){20}

111: IHLNL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

112: HLNLL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

113: LNLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

114: NLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAM96607 ck: 4751 len: 80 ! Aam96607 Human reproductive system related
(R,K){20}

61: KXXFD KKKKKKKKKKKKKKKKKKK
(K){20}

1 AAU18162 ck: 7907 len: 39 ! Aau18162 Novel human DNA-binding protein #9
(R,K){20}

9: YFEDL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: FEDLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: EDLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

12: DLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

13: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18167 ck: 9194 len: 87 ! Aau18167 Novel human DNA-binding protein #1
(R,K){20}

52: KIILL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

53: IILLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

54: ILLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

55: LLKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

56: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18168 ck: 8659 len: 104 ! Aau18168 Novel human DNA-binding protein #1
(R,K){20}

75: PLGGQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

76: LGGQK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

77: GGQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

78: GQKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

79: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18171 ck: 9398 len: 48 ! Aau18171 Novel human DNA-binding protein
(R,K){20}

2: Q KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

3: QK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

4: QKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18177 ck: 8278 len: 53 ! Aau18177 Novel human DNA-binding protein
(R,K){20}

30: NCGIL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

31: CGILK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

32: GILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

33: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

34: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18178 ck: 444 len: 66 ! Aau18178 Novel human DNA-binding protein
(R,K){20}

35: SMTFS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

36: MTFSS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18179 ck: 5503 len: 50 ! Aau18179 Novel human DNA-binding protein
(R,K){20}

30: IICLL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

31: ICILL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1 AAU18184 ck: 5691 len: 108 ! Aau18184 Novel human DNA-binding protein
(R,K){20}

78: VRPCL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

79: RPCLK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
80: PCLKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
81: CLKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
82: LKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
83: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
84: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
85: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
86: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKK
87: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKK
88: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} K
89: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK

1
AAU18192 ck: 6029 len: 63 1 Aau18192 Novel human DNA-binding protein #3
(R,K){20}
(K){20}
40: RLTL KKKKKKKKKKKKKKKKKKKKK ISWG

1
AAU18200 ck: 7170 len: 63 1 Aau18200 Novel human DNA-binding protein #4
(R,K){20}
(K){20}
37: TPSRA KKKKKKKKKKKKKKKKKKKKK KKKKK
38: PSRAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
39: SRAKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
40: RAKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
41: AKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKK
42: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KK
43: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} K
44: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK

1
AAU18204 ck: 6110 len: 61 1 Aau18204 Novel human DNA-binding protein #5
(R,K){20}

28: RPTRP KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
29: PTRPK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
30: TRPKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
31: RPKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
32: PKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
33: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
34: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
35: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
36: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
37: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK

1
AAU18205 ck: 5764 len: 74 1 Aau18205 Novel human DNA-binding protei
(R,K){20}
(K){20}
40: EPLSA KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
41: FLSAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
42: LSAAK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
43: SAAKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
44: AKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
45: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
46: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
47: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
48: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
49: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
50: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
51: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKKKK
52: KKKKK KKKKKKKKKKKKKKKKKKKKK (K){20} KKK

(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KX
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK X

AAU18206 ck: 9217 len: 68 1 Aau18206 Novel human DNA-binding protein #5

(R,K){20}

38: FLPEE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

39: LPPEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

40: FPEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

41: PEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

42: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK GXKF
(K){20}

AAU18208 ck: 8152 len: 79 1 Aau18208 Novel human DNA-binding protein #5

(R,K){20}

41: VRPEV RKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

42: RPRVR KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

43: PRVRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

44: RVRKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

45: VRKKK KKKKKKKKKKKKKKKKKKK GGRFR
(K){20}

AAU18237 ck: 285 len: 118 1 Aau18237 Novel human DNA-binding protein #8

(R,K){20}

98: EKHQK KKKKKKKKKKKKKKKKKKK G
(R,K){20}

AAU18238 ck: 5509 len: 58 1 Aau18238 Novel human DNA-binding protein #8

(R,K){20}

36: FYEVC KKKKKKKKKKKKKKKKKKK KKK
(K){20}

37: YFVCK KKKKKKKKKKKKKKKKKKK KK
(K){20}

38: FVCKK KKKKKKKKKKKKKKKKKKK K
(K){20}

(K){20}
39: VCKKK KKKKKKKKKKKKKKKKKKK

AAU18239 ck: 9074 len: 66 1 Aau18239 Novel human DNA-binding protein

(R,K){20}

40: LVQCE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

41: VQCEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

42: QCEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

43: CEKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

44: EKKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18240 ck: 8528 len: 150 1 Aau18240 Novel human DNA-binding protein

(R,K){20}

113: SRNTV KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

114: RNTVK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

115: NTVKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

116: TVKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

117: VRKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

118: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

119: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

120: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

121: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

122: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

123: KKKKK KKKKKKKKKKKKKKKKKKK XXAVL
(K){20}

AAU18241 ck: 7676 len: 156 1 Aau18241 Novel human DNA-binding protein

(R,K){20}

(K){20}

(K){20}

(K){20}

(K){20}

(K){20}

(K){20}

(K){20}

(K){20}

(K){20}

108: KTWI KKKKKKKKKKKKKKKKKKK
(K){20}
109: TWIK KKKKKKKKKKKKKKKKKKK
(K){20}
110: TWIK KKKKKKKKKKKKKKKKKKK
(K){20}
111: WIKK KKKKKKKKKKKKKKKKKKK
(K){20}
112: IKKK KKKKKKKKKKKKKKKKKKK
(K){20}
113: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
114: KKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}
115: KKKK KKKKKKKKKKKKKKKKKKK
(R,K){20}

AAU18242 ck: 1736 len: 40 1 Aau18242 Novel human DNA-binding protein #8

(R,K){20}
18: LPGL KKKKKKKKKKKKKKKKKKK
(K){20}
19: PGLK KKKKKKKKKKKKKKKKKKK
(K){20}
20: GLKK KKKKKKKKKKKKKKKKKKK
(K){20}
21: SLKK KKKKKKKKKKKKKKKKKKK
(K){20}

AAU18244 ck: 1109 len: 98 1 Aau18244 Novel human DNA-binding protein #9

(R,K){20}
53: QTKY KKKKKKKKKKKKKKKKKKK
(K){20}
54: TKTK KKKKKKKKKKKKKKKKKKK
(K){20}
55: KTKK KKKKKKKKKKKKKKKKKKK
(K){20}
56: NTKK KKKKKKKKKKKKKKKKKKK
(K){20}
57: TKKK KKKKKKKKKKKKKKKKKKK
(K){20}
58: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
59: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
60: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
61: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
62: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}

63: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
64: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
65: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
66: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
67: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
68: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}

AAU18246 ck: 8102 len: 111 1 Aau18246 Novel human DNA-binding protei

(R,K){20}
78: EFHL KKKKKKKKKKKKKKKKKKK
(K){20}
79: FHIL KKKKKKKKKKKKKKKKKKK
(K){20}
80: HILK KKKKKKKKKKKKKKKKKKK
(K){20}
81: ILKK KKKKKKKKKKKKKKKKKKK
(K){20}
82: LKKK KKKKKKKKKKKKKKKKKKK
(K){20}
83: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
84: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
85: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
86: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
87: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
88: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
89: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
90: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
91: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}
92: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}

AAU18247 ck: 8102 len: 111 1 Aau18247 Novel human DNA-binding protei

(R,K){20}
78: EFHL KKKKKKKKKKKKKKKKKKK
(K){20}

(K){20}
79: FHILK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
80: HILKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}
AAU18248 ck: 8319 len: 53 1 Aau18248 Novel human DNA-binding protein #9
(R,K){20}
(K){20}
13: RYFPP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: YFPPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: FPPKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KPPKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: PPKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAU18250 ck: 7918 len: 60 1 Aau18250 Novel human DNA-binding protein
(R,K){20}
(K){20}
50: NVLTV KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
51: VLTVK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
52: LTVKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: TVKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
AAU18252 ck: 4882 len: 41 1 Aau18252 Novel human DNA-binding protein
(R,K){20}
(K){20}
8: FYCFP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: YCFPP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: CFFPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: FFKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: FKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK GX
(K){20}

AAU18253 ck: 5469 len: 63 ! Aau18253 Novel human DNA-binding protein #1

(R,K){20}
30: ICCL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: ICCL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: CLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: LKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: LKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK XG
(K){20}

AAU18254 ck: 5075 len: 52 ! Aau18254 Novel human DNA-binding protein #1

(R,K){20}
30: FIVVX KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
31: IVVXK KKKKKKKKKKKKKKKKKKK KK
(K){20}
32: VVXKK KKKKKKKKKKKKKKKKKKK K
(K){20}
33: VXXKK KKKKKKKKKKKKKKKKKKK

AAU18255 ck: 5741 len: 47 ! Aau18255 Novel human DNA-binding prote

(R,K){20}
20: ILTF KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: LTFK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: TTFK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: TFKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: FKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK X
(K){20}

AAU18256 ck: 2868 len: 84 ! Aau18256 Novel human DNA-binding prote

(R,K){20}
53: KCTE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: CTYE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: TYEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: YEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: EKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAU18257 ck: 4686 len: 73 ! Aau18257 Novel human DNA-binding prote

1

(R,K){20}

(K){20}

41: YLKEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

42: LKKEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

43: KKEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

44: KEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

45: EKKEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18258 ck: 6676 len: 74 1 Aau18258 Novel human DNA-binding protein #1

(R,K){20}

(K){20}

47: LPTPQ KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: RTPQK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: TFOKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: FQKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

AAU18259 ck: 2283 len: 54 1 Aau18259 Novel human DNA-binding protein #1

(R,K){20}

(K){20}

32: TVPCF KKKKKKKKKKKKKKKKKKK KKK

1

(R,K){20}

(K){20}

AAU18260 ck: 7503 len: 74 1 Aau18260 Novel human DNA-binding protein

33: VFCFK KKKKKKKKKKKKKKKKKKK KX

(K){20}

34: FCFKK KKKKKKKKKKKKKKKKKKK X

45: SHLTD KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

46: HLTDK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

47: LTDKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: TDKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

49: DKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

52: KKKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

53: KKKKK KKKKKKKKKKKKKKKKKKK KK

(K){20}

54: KKKKK KKKKKKKKKKKKKKKKKKK K

(K){20}

55: KKKKK KKKKKKKKKKKKKKKKKKK

AAU18262 ck: 5199 len: 84 1 Aau18262 Novel human DNA-binding protein

(R,K){20}

(K){20}

63: AMNAS KKKKKKKKKKKKKKKKKKK XG

AAU18263 ck: 7578 len: 31 1 Aau18263 Novel human DNA-binding protein

(R,K){20}

(K){20}

6: LTELE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

7: TELEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

8: ELEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

9: LEKKK KKKKKKKKKKKKKKKKKKK KKK

(K){20}

10: EKKKK KKKKKKKKKKKKKKKKKKK KX

(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKK X

```
1
AAU18264 ck: 3915 len: 57 ! Aau18264 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
31: KQLLL KKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
32: QLLK KKKKKKKKKKKKKKKKKKK KXXG
(K){20}
33: LLLK KKKKKKKKKKKKKKKKKKK KXGGE
(K){20}
34: LLKK KKKKKKKKKKKKKKKKKKK XGGF
(K){20}

1
AAU18265 ck: 3679 len: 37 ! Aau18265 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
15: ISPL KKKKKKKKKKKKKKKKKKK KXX
(K){20}
16: SPLK KKKKKKKKKKKKKKKKKKK KX
(K){20}
17: PLTK KKKKKKKKKKKKKKKKKKK X
(K){20}

1
AAU18266 ck: 657 len: 196 ! Aau18266 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
169: FYFE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
170: VYFE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
171: XFEK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
172: FEKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
173: EKKK KKKKKKKKKKKKKKKKKKK KKKX
(K){20}
174: KKKK KKKKKKKKKKKKKKKKKKK KXX
(K){20}
175: KKKK KKKKKKKKKKKKKKKKKKK KX
(K){20}
176: KKKK KKKKKKKKKKKKKKKKKKK X
(K){20}

1
AAU18267 ck: 4672 len: 57 ! Aau18267 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
28: DTFE KKKKKKKKKKKKKKKKKKK KKKKX
(K){20}
29: KTFE KKKKKKKKKKKKKKKKKKK KKKXP
(K){20}
30: TFEK KKKKKKKKKKKKKKKKKKK KXXPG
(K){20}
31: FEKK KKKKKKKKKKKKKKKKKKK KXXPG
(K){20}

1
32: HKKK KKKKKKKKKKKKKKKKKKK KXXGG
(K){20}
AAU18268 ck: 9656 len: 66 ! Aau18268 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
38: WVSV KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
39: VVSV KKKKKKKKKKKKKKKKKKK KKKRE
(K){20}
40: VVSV KKKKKKKKKKKKKKKKKKK KKKRE
(K){20}
41: SVKK KKKKKKKKKKKKKKKKKKK KKKEX
(K){20}
42: VKKK KKKKKKKKKKKKKKKKKKK KKKXX
(K){20}
43: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

1
AAU18270 ck: 4665 len: 34 ! Aau18270 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
10: PELL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
11: ELLK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
12: LLKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
13: LLKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
14: LKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
15: KKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

1
AAU18271 ck: 7810 len: 64 ! Aau18271 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
37: LKFE KKKKKKKKKKKKKKKKKKK KKKKX
(K){20}
38: KYFE KKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
39: YFEK KKKKKKKKKKKKKKKKKKK KKKXG
(K){20}
40: FEKK KKKKKKKKKKKKKKKKKKK KXXGP
(K){20}
41: WKKK KKKKKKKKKKKKKKKKKKK XGXP
(K){20}

1
AAU18272 ck: 269 len: 76 ! Aau18272 Novel human DNA-binding protein #1
(R,K){20}
(K){20}
37: TFSR KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
```

38: PSRAK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
39: SPRAK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
40: RAKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
41: AKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

AAU18273 ck: 8370 len: 45 1 Aau18273 Novel human DNA-binding protein #1
(R,K){20}
17: APRTQ KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
18: PRTQK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
19: KTOKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
20: TOKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
21: OKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

AAU18274 ck: 1663 len: 87 1 Aau18274 Novel human DNA-binding protein #1
(R,K){20}
36: KMSXX KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
37: WSSXX KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
38: SSXXX KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
39: SXKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
40: XKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}

AAU18275 ck: 5607 len: 63 1 Aau18275 Novel human DNA-binding protein
(R,K){20}
26: MVELE KKKKKKKKKKKKKKKKKKK KKKK
(K){20}


```

27: VLEKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
28: ELEKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
29: LEKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
30: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG      (K){20}
      (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG      (K){20}
      (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGP      (K){20}
      (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK GGPF      (K){20}
      (K){20}

AAU18276   ck: 5997   len: 58    ! Aau18276 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
28: RPTRP KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
29: PTPPK KKKKKKKKKKKKKKKKKKKK XXXKK      (K){20}
      (K){20}

AAU18277   ck: 5764   len: 74    ! Aau18277 Novel human DNA-binding protein #1
      (R,K){20}
      (K){20}
40: EFLSA KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
41: FLSSK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
42: LSARK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
43: SAKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
44: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
      (K){20}

```

```

46: KKKK (K){20} KKKK
47: KKKK (K){20} KKKK
48: KKKK (K){20} KKKK
49: KKKK (K){20} KKKK
50: KKKK (K){20} KKKK
51: KKKK (K){20} KKKK
52: KKKK (K){20} KKKK
53: KKKK (K){20} KKKK
54: KKKK (K){20} KKKK
55: KKKK (K){20} KKKK
56: KKKK (K){20} KKKK
57: KKKK (K){20} KKKK
58: KKKK (K){20} KKKK
59: KKKK (K){20} KKKK
60: KKKK (K){20} KKKK
61: KKKK (K){20} KKKK
62: KKKK (K){20} KKKK
63: KKKK (K){20} KKKK
64: KKKK (K){20} KKKK
65: KKKK (K){20} KKKK
66: KKKK (K){20} KKKK
67: KKKK (K){20} KKKK
68: KKKK (K){20} KKKK
69: KKKK (K){20} KKKK
70: KKKK (K){20} KKKK

```

71: KKKKK (K){20} KKKKK
 72: KKKKK (K){20} KKKKK
 73: KKKKK (K){20} KKKKK
 74: KKKKK (K){20} KKKKK
 75: KKKKK (K){20} KKKKK

AAU18279 ck: 3164 len: 181 1 Aau18279 Novel human DNA-binding protein #1
 154: TRKPE (R,K){20} KKKKK
 155: RKPEK (K){20} KKKKK
 156: KPEKK (K){20} KKKKK
 157: PEKKK (K){20} KKKKK
 158: EKKKK (K){20} KKKKK
 159: KKKKK (K){20} KKKKK
 160: KKKKK (K){20} KKKKK
 161: KKKKK (K){20} KKKKK

AAU18280 ck: 7117 len: 55 1 Aau18280 Novel human DNA-binding protein #1
 22: DDKKN (R,K){20} KKKKK
 23: DKKNK (K){20} KKKKK
 24: KKNKK (K){20} KKKKK
 25: KNNKK (K){20} KKKKK
 26: NKKKK (K){20} KKKKK
 27: KKKKK (K){20} KKKKK

AAU18281 ck: 9316 len: 67 1 Aau18281 Novel human DNA-binding protein #1
 36: FTLOT (R,K){20} KKKKK

37: TLQTR (K){20} KKKKK
 38: LOTRK (K){20} KKKKK
 39: QTRKK (K){20} KKKKK
 40: TRKKK (K){20} KKKKK
 41: RKKKK (K){20} KKKKK

AAU53862 ck: 5383 len: 86 1 Aau53862 Human brain expressed single ex
 15: RRRNG (R,K){20} KRRRR
 16: RRGRK (R,K){20} KRRRR
 17: RRRGR (R,K){20} KRRRR
 18: RGRRR (R,K){20} KRRRR
 19: GRRRR (R,K){20} KRRRR
 20: RRRRR (R,K){20} KRRRR
 21: RRRRK (R,K){20} KRRRR
 22: RRRKK (R,K){20} KRRRR
 23: RRRKK (R,K){20} KRRRR
 24: RRRKK (R,K){20} KRRRR
 25: KRRKK (R,K){20} KRRRR
 26: KRRKK (R,K){20} KRRRR
 27: KRRKR (R,K){20} KRRRR
 28: KRRKK (R,K){20} KRRRR
 29: KRRKK (R,K){20} KRRRR
 30: KRRKK (R,K){20} KRRRR
 31: RRRKK (R,K){20} KRRRR
 32: KRRKK (R,K){20} KRRRR

AA054700 ck: 1334 len: 86 ! Aam54700 Human brain expressed single exon

```
(R,K){20}
33: KKKK KKKRRRRRRRRRRRR RR
(R,K){20}
34: KKKK KKKRRRRRRRRRRRR RR
(R,K){20}
35: KKKK KKKRRRRRRRRRRRR RR
(R,K){20}
36: KKKK KKKRRRRRRRRRRRR RR
(R,K){20}
37: KKKK RRRRRRRRRRRRRRR RR
(R){20}
38: KKKK RRRRRRRRRRRRRRR RR
(R){20}
39: KKKK RRRRRRRRRRRRRRR RR
(R){20}
40: KKKK RRRRRRRRRRRRRRR RR
(R){20}
41: KKKK RRRRRRRRRRRRRRR RR
(R){20}
42: KKKK RRRRRRRRRRRRRRR RR
(R){20}
43: KKKK RRRRRRRRRRRRRRR RR
(R){20}
44: KKKK RRRRRRRRRRRRRRR RR
(R){20}
45: KKKK RRRRRRRRRRRRRRR RR
(R){20}
(R,K){20}
(R,K){20}
57: EEEG RRRRRRRRRRRRR RR
(R,K){20}
58: EEGR RRRRRRRRRRRRR RR
(R,K){20}
59: EEGR RRRRRRRRRRRRR RR
(R,K){20}
60: EGRR KKKRRRRRRRRRR RR
(R,K){20}
61: GRRK KKKRRRRRRRRRR RR
(R,K){20}
62: RRRK KKKRRRRRRRRRR RR
(R,K){20}
63: RRRK KKKRRRRRRRRRR RR
(R,K){20}
64: RKKK RRRRRRRRRRRRR RR
(R,K){20}
65: KKKR KKKRRRRRRRRRR RR
(R,K){20}
66: KKKR RRRRRRRRRRRRR RR
```

```

1
67: KKKR KKKKKKKKKKKKKKKKKKK
(A,K){20}
AAM54795 ck: 9082 len: 167 ! Aam54795 Human brain expressed single e:
(R,K){20}
33: EEGRG RRRRRRRKKRRKKRRKKR RRGCG
(R,K){20}
34: EGRGR RRRRRRRKKRRKKRRKKR RGGGR
(R,K){20}
35: GGGRR RRRRRRRKKRRKKRRKKR GGGRR
(R,K){20}
AAM55320 ck: 2276 len: 89 ! Aam55320 Human brain expressed single e:
(R,K){20}
23: EEEEE KKKKKRRKKKKKKKKKKK EEEKK
(K){20}
46: KKEEE KKKKKKKKKKKKKKKK KRKEE
(K){20}
47: KEEER KKKKKKKKKKKKKKKK RKEEE
(R,K){20}
48: EEEKK KKKKKKKKKKKKKKKK KEEEE
(R,K){20}
49: EEEKK KKKKKKKKKKKKKKKK EEEEE
(R,K){20}
AAM55623 ck: 1939 len: 130 ! Aam55623 Human brain expressed single e:
(R,K){20}
42: EGRRK RRRRRRRRRRRRRRRKK RRRR
(R,K){20}
43: GRKER RRRRRRRRRRRRRKK RRRR
(R,K){20}
44: KEERR RRRRRRRRRRRRRKK RRRK
(R,K){20}
45: KEERR RRRRRRRRRRRRRKK RRRK
(R,K){20}
46: EERRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
47: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
48: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
49: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
50: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
51: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}
52: RRRRR RRRRRRRRRRRRRKK RRRK
(R,K){20}

```

53: RRRR RRRRRRRRRRRRRRRRRR KEEE
(R,K){20}
54: RRRR RRRRRRRRRRRRRRRRRR EEEE
AAM5931 ck: 3607 len: 88 1 Aam5931 Human brain expressed single exon
(R,K){20}
39: ERRE RRRRRRRRRRRRRRRRRR KKKK
(K){20}
40: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
41: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
42: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
43: ERRE RRRRRRRRRRRRRRRRRR KKKK
(K){20}
44: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
45: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
46: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
47: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
48: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
49: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
50: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
51: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
52: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
53: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
54: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
55: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
56: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
57: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
58: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
59: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}

60: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
61: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
62: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
63: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
64: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
65: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
66: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
67: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
68: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
69: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
AAM56490 ck: 3937 len: 85 1 Aam56490 Human brain expressed single exon
(R,K){20}
1: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
2: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
3: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
4: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
5: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
6: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
7: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
8: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
9: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
10: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
11: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}
AAM58224 ck: 1560 len: 88 1 Aam58224 Human brain expressed single exon
(R,K){20}
43: RRRR RRRRRRRRRRRRRRRRRR KKKK
(K){20}

1

6: ETERE KKKKKKKKKKKKKKKKKKKK KKKKK

AAM61063 ck: 8343 len: 66 ; Aam61063 Human brain expressed single exon

1

16: KERKT KKKKKKKKKKKKKKKKKR R
17: ERKTK KKKKKKKKKKKKKKKKKRR

AAM58395 ck: 2324 len: 36 ; Aam58395 Human brain expressed single exon

62: RRRR RRRRRRRRRRRRRRRR NNNE

61: RRRR RRRRRRRRRRRRRRRR RNTNN
60: RRRR RRRRRRRRRRRRRRRR RNTN
59: RRRR RRRRRRRRRRRRRRRR RRRT
58: RRRR RRRRRRRRRRRRRRRR RRRN
57: RRRR RRRRRRRRRRRRRRRR RRRR
56: RRRR RRRRRRRRRRRRRRRR RRRR
55: RRRR RRRRRRRRRRRRRRRR RRRR
54: RRRR RRRRRRRRRRRRRRRR RRRR
53: RRRR RRRRRRRRRRRRRRRR RRRR
52: RRRR RRRRRRRRRRRRRRRR RRRR
51: RRRR RRRRRRRRRRRRRRRR RRRR
50: RRRR RRRRRRRRRRRRRRRR RRRR
49: RRRR RRRRRRRRRRRRRRRR RRRR
48: RRRR RRRRRRRRRRRRRRRR RRRR
47: GRRR RRRRRRRRRRRRRRRR RRRR
46: GRRR RRRRRRRRRRRRRRRR RRRR
45: RGRH RRRRRRRRRRRRRRRR RRRR
44: ERGR RRRRRRRRRRRRRRRR RRRR

```

(R,K){20}
7: TEREK KKKKKKKKKKKKKKKKKK KKKK
8: EREKK KKKKKKKKKKKKKKKKKK KKKK
(R,K){20}
9: REKK KKKKKKKKKKKKKKKKK NKKK
AAM6353 CK: 2394 Len: 57 | Aam6353 Human brain expressed single e
(R,K){20}
20: EEEEG RRRRRRRRRRRRRRRR RRRR
(R,K){20}
21: EEGR RRRRRRRRRRRRRR RRGG
(R,K){20}
22: EEGR RRRRRRRRRRRRRR RRRG
(R,K){20}
23: EGRR RRRRRRRRRRRRRR RGRR
(R,K){20}
24: GRRR RRRRRRRRRRRRRR RGRR
(R,K){20}
25: RRRR RRRRRRRRRRRRRR GRRR
AAM64090 CK: 4228 Len: 24 | Aam64090 Human brain expressed single e
(R,K){20}
1: RRRRRRRRRRRRRRRR RRTT
(R,K){20}
2: R RRRRRRRRRRRRRR RRT
(R,K){20}
3: RR RRRRRRRRRRRRRR RT
(R,K){20}
4: RRR RRRRRRRRRRRRRR T
AAM64863 CK: 3301 Len: 52 | Aam64863 Human brain expressed single e
(R,K){20}
12: KKKKN KKKKKKKKKKKKKKK KKKK
(R,K){20}
13: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
14: KKNN KKKKKKKKKKKKKKK KKKK
(R,K){20}
15: KNKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
16: NKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
17: KKKK KKKKKKKKKKKKKKK KKKK
(R,K){20}
18: KKKK KKKKKKKKKKKKKKK KKKK

```

```

19: KKKKK (R,K){20}
20: KKKKK (R,K){20}
21: KKKKK (R,K){20}
22: KKKKK (R,K){20}
23: KKKKK (R,K){20}
24: KKKKK (R,K){20}
25: KKKKK (R,K){20}
26: KKKKK (R,K){20}
27: KKKKK (R,K){20}
28: KKKKK (R,K){20}
29: KKKKK (R,K){20}
30: KKKKK (R,K){20}
31: KKKKK (R,K){20}

AAM65355 ck: 4895 len: 51 | Aam65355 Human brain expressed single exon
18: LKKKK (R,K){20}
19: KKKKK (R,K){20}
20: KKKKK (R,K){20}

AAM66249 ck: 5383 len: 86 | Aam66249 Human bone marrow expressed probe
15: KKKKK (R,K){20}
16: KKKKK (R,K){20}
17: KKKKK (R,K){20}
18: KKKKK (R,K){20}
19: KKKKK (R,K){20}
20: KKKKK (R,K){20}

```

```

21: KKKKK (R,K){20}
22: KKKKK (R,K){20}
23: KKKKK (R,K){20}
24: KKKKK (R,K){20}
25: KKKKK (R,K){20}
26: KKKKK (R,K){20}
27: KKKKK (R,K){20}
28: KKKKK (R,K){20}
29: KKKKK (R,K){20}
30: KKKKK (R,K){20}
31: KKKKK (R,K){20}
32: KKKKK (R,K){20}
33: KKKKK (R,K){20}
34: KKKKK (R,K){20}
35: KKKKK (R,K){20}
36: KKKKK (R,K){20}
37: KKKKK (R,K){20}
38: KKKKK (R,K){20}
39: KKKKK (R,K){20}
40: KKKKK (R,K){20}
41: KKKKK (R,K){20}
42: KKKKK (R,K){20}
43: KKKKK (R,K){20}
44: KKKKK (R,K){20}

```

45: RRRR (R){20}
RRRRRRRRRRRRRRRRRRR NKOTK

AAm67100 ck: 1334 len: 86 ! Aam67100 Human bone marrow expressed probe

57: EEEG (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

58: EEEG (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

59: EGGG (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

60: EGGG (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

61: GRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

62: RRRK (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKKK

63: RRRK (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKK

64: RRRK (R,K){20}
RRRRRRRRRRRRRRRRRRR KKK

65: KKKR (R,K){20}
RRRRRRRRRRRRRRRRRRR K

66: KKKR (R,K){20}
RRRRRRRRRRRRRRRRRRR K

67: KKKR (K){20}
RRRRRRRRRRRRRRRRRRR

AAm67180 ck: 9082 len: 167 ! Aam67180 Human bone marrow expressed probe

33: EGGG (R,K){20}
RRRRRRRRRRRRRRRRRRR RGGG

34: EGGG (R,K){20}
RRRRRRRRRRRRRRRRRRR RGGG

35: GGGG (R,K){20}
RRRRRRRRRRRRRRRRRRR GGGG

AAm67717 ck: 2276 len: 89 ! Aam67717 Human bone marrow expressed probe

23: EEEE (R,K){20}
RRRRRRRRRRRRRRRRRRR EEEK

46: KKEE (K){20}
RRRRRRRRRRRRRRRRRRR KKEE

47: KKEE (K){20}
RRRRRRRRRRRRRRRRRRR KKEE

48: EEEK (R,K){20}
RRRRRRRRRRRRRRRRRRR EEEK

49: EEEK (R,K){20}
RRRRRRRRRRRRRRRRRRR EEEK

1

AAm68007 ck: 1939 len: 130 ! Aam68007 Human bone marrow expressed pr

42: EGKE (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

43: EGKE (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

44: RKEE (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

45: KERR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

46: EERR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

47: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

48: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

49: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

50: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

51: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

52: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

53: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

54: RRRR (R,K){20}
RRRRRRRRRRRRRRRRRRR RRRR

AAm68298 ck: 3607 len: 88 ! Aam68298 Human bone marrow expressed pr

39: EKKR (R,K){20}
RRRRRRRRRRRRRRRRRRR KKKK

40: RKEE (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

41: KKEE (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

42: RKEE (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

43: EKKK (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

44: KKKK (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

45: KKKK (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

46: KKKK (K){20}
RRRRRRRRRRRRRRRRRRR KKKK

47: KKKK (R){20}
RRRRRRRRRRRRRRRRRRR KKKK

48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
50: KKKKK (K){20} KKKKK
51: KKKKK (K){20} KKKKK
52: KKKKK (K){20} KKKKK
53: KKKKK (K){20} KKKKK
54: KKKKK (K){20} KKKKK
55: KKKKK (K){20} KKKKK
56: KKKKK (K){20} KKKKK
57: KKKKK (K){20} KKKKK
58: KKKKK (K){20} KKKKK
59: KKKKK (K){20} KKKKK
60: KKKKK (K){20} KKKKK
61: KKKKK (K){20} KKKKK
62: KKKKK (K){20} KKKKK
63: KKKKK (K){20} KKKKK
64: KKKKK (K){20} KKKKK
65: KKKKK (K){20} KKKKK
66: KKKKK (K){20} KKKKK
67: KKKKK (K){20} KKKKK
68: KKKKK (K){20} KKKKK
69: KKKKK (K){20} KKKKK
AAM68869 ck: 3937 len: 85 1 Aam68869 Human bone marrow expressed probe
(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK

2: K (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
3: KK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
4: KKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
5: KKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
7: KKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
8: KKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
9: KKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
10: KKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
11: KKKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
AAM69569 ck: 2686 len: 71 1 Aam69569 Human bone marrow expressed probe
(R,K){20}
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
25: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
26: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
27: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
28: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
29: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
30: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
31: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK
32: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKKKKKKKKKK


```
(R,K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM70678 ck: 1560 len: 88 1 Aam70678 Human bone marrow expressed probe
(R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM70678 Human bone marrow expressed probe
```

```
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM70861 ck: 2324 len: 36 1 Aam70861 Human bone marrow expressed probe
(R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM73767 ck: 8343 len: 66 1 Aam73767 Human bone marrow expressed probe
(R,K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
AAM76347 ck: 2394 len: 57 1 Aam76347 Human bone marrow expressed probe
(R,K){20}
```

20: EEEEG RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
21: EEEGR RRRRRRRRRRRRRRRRRR RRRG
(R,K){20}
22: EEEGR RRRRRRRRRRRRRRRRRR RRRG
(R,K){20}
23: EGGRR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
24: GRRRR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
25: RRRRR RRRRRRRRRRRRRRRRRR RRRR

AAM76911 ck: 4228 len: 24 ! Aam76911 Human bone marrow expressed probe

1: (R,K){20}
(R,K){20}
RRRRRRRRRRRRRRRRRR RRRR
2: R RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
3: RR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
4: RRR RRRRRRRRRRRRRRRRR RRRR

AAM78048 ck: 4895 len: 51 ! Aam78048 Human bone marrow expressed probe

1: (R,K){20}
(R,K){20}
18: LKPPM RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
19: FKPMR RRRRRRRRRRRRRRRRR RRRR
(R,K){20}
20: KPMRK RRRRRRRRRRRRRRRRR RRRR

AAM82533 ck: 1736 len: 40 ! Aam82533 Human immune/haematopoietic antigen

1: (R,K){20}
(K){20}
18: LPSGL RRRRRRRRRRRRRRRRR RRRR
(K){20}
19: PGSGL RRRRRRRRRRRRRRRRR RRRR
(K){20}
20: GSLKK RRRRRRRRRRRRRRRRR RRRR
(K){20}
21: SLKKK RRRRRRRRRRRRRRRRR RRRR

AAM85748 ck: 7503 len: 74 ! Aam85748 Human immune/haematopoietic antigen

1: (R,K){20}
(K){20}
45: SHLTD RRRRRRRRRRRRRRRRR RRRR
(K){20}
46: HLTDR RRRRRRRRRRRRRRRRR RRRR
(K){20}

47: LTDRK RRRRRRRRRRRRRRRRR RRRR

48: TDRKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

49: DRKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

50: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

51: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

52: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

53: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

54: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

55: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

AAM90546 ck: 6676 len: 74 ! Aam90546 Human immune/haematopoietic antigen

1: (R,K){20}
(K){20}

47: LRTFQ RRRRRRRRRRRRRRRRR RRRR
(K){20}

48: RTFOK RRRRRRRRRRRRRRRRR RRRR
(K){20}

49: TFOKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

50: FOQKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

51: QKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

52: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

AAM90618 ck: 5691 len: 108 ! Aam90618 Human immune/haematopoietic antigen

1: (R,K){20}
(K){20}

78: VRPCL RRRRRRRRRRRRRRRRR RRRR
(K){20}

79: RPCLK RRRRRRRRRRRRRRRRR RRRR
(K){20}

80: PCLKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

81: CLKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

82: LKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

83: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

84: KKKKK RRRRRRRRRRRRRRRRR RRRR
(K){20}

```

85: KKKK KKKKKKKKKKKKKKKKK KKK      (K){20}
86: KKKK KKKKKKKKKKKKKKKKK KK        (K){20}
87: KKKK KKKKKKKKKKKKKKKKK KK        (K){20}
88: KKKK KKKKKKKKKKKKKKKKK K         (K){20}
89: KKKK KKKKKKKKKKKKKKKKK K         (K){20}
90: KKKK KKKKKKKKKKKKKKKKK KK        (K){20}
91: KKKK KKKKKKKKKKKKKKKKK K         (K){20}
92: KKKK KKKKKKKKKKKKKKKKK          (K){20}

AAM91030 ck: 9102 len: 111 ! Aam91030 Human immune/haematopoietic antigen
(R,K){20}
(K){20}
78: EFHL KKKKKKKKKKKKKKKKK KKKK     (R,K){20}
79: FHLK KKKKKKKKKKKKKKKKK KKKK     (K){20}
80: HLK KKKKKKKKKKKKKKKKK KKKK       (K){20}
81: ILK KKKKKKKKKKKKKKKKK KKKK       (K){20}
82: LKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
83: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
84: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
85: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
86: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
87: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
88: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
89: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
90: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
91: KKK KKKKKKKKKKKKKKKKK KKKK       (K){20}
92: KKK KKKKKKKKKKKKKKKKK          (K){20}

AAW91162 ck: 1109 len: 98 ! Aaw91162 Human immune/haematopoietic antigen
(R,K){20}
(K){20}
53: OTKT KKKKKKKKKKKKKKKKK KKKK     (K){20}
54: TKT KKKKKKKKKKKKKKKKK KKKK       (K){20}

```

```

55: KNTKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
56: NTKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
57: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (X){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (X){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (X){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRS      (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK GGRSR      (K){20}

AAW91891 ck: 8102 len: 111 ! Aaw91891 Human immune/hematopoietic anti
          (R,K){20}
          (K){20}
78: EFHLL KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
79: FHILL KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK      (K){20}

```

87: KKKK (K){20} KKKK
88: KKKK (K){20} KKKK
89: KKKK (K){20} KKKK
90: KKKK (K){20} KKKK
91: KKKK (K){20} K
92: KKKK (K){20} KKKK
AAM9193 ck: 8285 len: 55 ! Aam9193 Human digestive system antigen SEQ
(R,K){20}
33: PTPRP KKKK (K){20} KKKK
34: PTPRK KKKK (K){20} GK
AAM92433 ck: 1663 len: 87 ! Aam92433 Human digestive system antigen SEQ
(R,K){20}
36: KWSX KKKK (K){20} KKKK
37: WSSXK KKKK (K){20} KKKK
38: SSXK KKKK (K){20} KKKK
39: SXKK KKKK (K){20} KKKK
40: XKKK KKKK (R){20} KKKK
41: KKKK KKKK (K){20} KKKK
42: KKKK KKKK (R){20} KKKK
43: KKKK KKKK (K){20} KKKK
44: KKKK KKKK (K){20} KKKK
45: KKKK KKKK (K){20} KKKK
46: KKKK KKKK (K){20} KKKK
47: KKKK KKKK (K){20} KKKK
48: KKKK KKKK (K){20} KKKK
(K){20}

49: KKKK KKKK
50: KKKK (K){20} KKKK
51: KKKK (K){20} KKKK
52: KKKK (K){20} KKKK
53: KKKK (K){20} KKKK
54: KKKK (K){20} KKKK
55: KKKK (K){20} KKKK
56: KKKK (K){20} KKKK
57: KKKK (K){20} KKKK
58: KKKK (K){20} KKKK
59: KKKK (K){20} KKKK
60: KKKK (K){20} KKKK
61: KKKK (K){20} KKKK
62: KKKK (K){20} KKKK
63: KKKK (R,K){20} KKKK
AaO0092 ck: 9065 len: 113 ! AaO0092 Human polypeptide SEQ ID NO 139
(R,K){20}
91: CLGCL KKKK (K){20} KKF
92: LGCLK KKKK (K){20} FF
AaO00222 ck: 2916 len: 132 ! AaO00222 Human polypeptide SEQ ID NO 141
(R,K){20}
28: XPLPP KKKK (K){20} GPPK
AaO00232 ck: 1000 len: 102 ! AaO00232 Human polypeptide SEQ ID NO 141
(R,K){20}
29: RDCFP KKKK (K){20} KKKK
30: DCFEK KKKK (K){20} KKKK
31: CFEKK KKKK (K){20} KKKK

32: FFKKK (R){20} KKKKKKKKKKKKKKKKKKK KKKKK
33: FKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
34: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
43: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
44: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
45: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
46: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
54: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

1
AA000291 ck: 8100 len: 124 1 Aa000291 Human polypeptide SEQ ID NO 141
(R,K){20}
22: QHFCM KKKKKKKKKKKKKKKKKKK FFKKK
(K){20}
1
AA000439 ck: 6396 len: 122 1 Aa000439 Human polypeptide SEQ ID NO 141
(R,K){20}
23: CLMLV KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: LMLVK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: WLVRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: LVKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: VKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
32: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
33: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
34: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
41: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
42: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
43: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK


```
(K)[20]
51: KKKK KKKKKKKKKKKKKKK KKKO
(K)[20]
52: KKKK KKKKKKKKKKKKKKK KKO
(K)[20]
53: KKKK KKKKKKKKKKKKKKK KQ
(K)[20]
54: KKKK KKKKKKKKKKKKKKK Q

AA002186 ck: 4844 len: 57 1 Aa002186 Human polypeptide SEQ ID NO 16078.
(R,K)[20]
11: HCCLL KKKKKKKKKKKKKKK KKKK
(K)[20]
12: CCLLK KKKKKKKKKKKKKKK IKKK

AA002310 ck: 704 len: 137 1 Aa002310 Human polypeptide SEQ ID NO 16202.
(R,K)[20]
22: HSLNL KKKKKKKKKKKKKKKR GGVK

AA002477 ck: 2018 len: 112 1 Aa002477 Human polypeptide SEQ ID NO 16369.
(R,K)[20]
83: LASAV KKKKKKKKKKKKKKK KKKK
(K)[20]
84: ASAVK KKKKKKKKKKKKKKK KKKK
(K)[20]
85: SAVKK KKKKKKKKKKKKKKK KKKK
(K)[20]
86: AVKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
87: VKKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
88: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
89: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
90: KKKK KKKKKKKKKKKKKKK KKK
(K)[20]
91: KKKK KKKKKKKKKKKKKKK KK
(K)[20]
92: KKKK KKKKKKKKKKKKKKK K
(K)[20]
93: KKKK KKKKKKKKKKKKKKK

AA002733 ck: 2860 len: 132 1 Aa002733 Human polypeptide SEQ ID NO 16625.
(R,K)[20]
86: FFFSL KKKKKKKKKKKKKKK KKKK
```

```
(K)[20]
87: FFSLK KKKKKKKKKKKKKKK KKKK
(K)[20]
88: FSLKK KKKKKKKKKKKKKKK KKKK
(K)[20]
89: SLKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
90: LKKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
91: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
92: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
93: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
94: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
95: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
96: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
97: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
98: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
99: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
100: KKKK KKKKKKKKKKKKKKK KKKK
(R,K)[20]
101: KKKK KKKKKKKKKKKKKKKR AASRS

AA002946 ck: 3649 len: 126 1 Aa002946 Human polypeptide SEQ ID NO 16
(R,K)[20]
25: DEATS KKKKKKKKKKKKKKK KKKK
(K)[20]
26: EATSK KKKKKKKKKKKKKKK KKKK
(K)[20]
27: ATSKK KKKKKKKKKKKKKKK KKKK
(K)[20]
28: TSKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
29: SKKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
30: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
31: KKKK KKKKKKKKKKKKKKK KKKK
(K)[20]
32: KKKK KKKKKKKKKKKKKKK KKKK
```

```
1
33: KKKKK (K){20} KKKKK
34: KKKKK (K){20} KKKKK
35: KKKKK (K){20} KKKKK
36: KKKKK (K){20} KKKKK
37: KKKKK (K){20} KGGGP
38: KKKKK (K){20} GGGFK

AAO02961 ck: 4320 len: 83 ! Aao02961 Human polypeptide SEQ ID NO 16853
(R,K){20}
(K){20}
22: RMFSS KKKKKKKKKKKKKKKKKKK KTAIT
(K){20}
23: MESSK KKKKKKKKKKKKKKKKKKK TAITK
(K){20}

AAO03006 ck: 5212 len: 102 ! Aao03006 Human polypeptide SEQ ID NO 16898
(R,K){20}
(K){20}
59: HDEFP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: DFEFP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: FEPFK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: FPKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
63: PKKKK KKKKKKKKKKKKKKKKKKK KKKKA
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKAS
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKK KKAAS
(K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKK KASSS
(K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKK ASSSS
(K){20}

AAO03024 ck: 6865 len: 64 ! Aao03024 Human polypeptide SEQ ID NO 16916
(R,K){20}
(R,K){20}
15: SPAKA KKKKKKKKKKKKKKKKKKK KRGGP
(K){20}
16: PAKAR KKKKKKKKKKKKKKKKKKK RGGPP
(R,K){20}
17: AKARK KKKKKKKKKKKKKKKKKKK GGPPL
(K){20}
```

```
1
AAO03113 ck: 5587 len: 60 ! Aao03113 Human polypeptide SEQ ID NO 170
(R,K){20}
(K){20}
33: EINML KKKKKKKKKKKKKKKKKKK KKKRG
(K){20}
34: INMLK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}
35: NMLKK KKKKKKKKKKKKKKKKKKK KRGGG
(K){20}
36: MLKKK KKKKKKKKKKKKKKKKKKK RGGGL
(R,K){20}
37: LKKKK KKKKKKKKKKKKKKKKKKK GGGL

AAO03123 ck: 6627 len: 108 ! Aao03123 Human polypeptide SEQ ID NO 170
(R,K){20}
(K){20}
33: RTWRX KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: TWRXX KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: WRXXK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: RXKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: XXKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKRG
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KRGGG
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK RGGGO
(R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK GGGOK
(K){20}

AAO03132 ck: 3903 len: 116 ! Aao03132 Human polypeptide SEQ ID NO 170
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKK KKKRG
(K){20}
3: KK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKK KRGGP
(K){20}
```


5: KKKK KKKKKKKKKKKKKKKKKR RGPP
(R,K){20}
6: KKKKK KKKKKKKKKKKKKKKKKR GGPRK
AAO03152 ck: 7891 len: 35 1 Aao03152 Human polypeptide SEQ ID NO 17044
(R,K){20}
13: LIXYT RKKKKKKKKKKKKKKKKK RGG
(R,K){20}
14: IXYTR KKKKKKKKKKKKKKKKKR GG
AAO03168 ck: 7233 len: 67 1 Aao03168 Human polypeptide SEQ ID NO 17060
(R,K){20}
17: ELIMP KKKKKKKKKKKKKKKKKK KRGA
(K){20}
18: LMPK KKKKKKKKKKKKKKKKKK RGAP
(K){20}
19: LMPK KKKKKKKKKKKKKKKKKR GGAPF
(R,K){20}
AAO03243 ck: 411 len: 95 1 Aao03243 Human polypeptide SEQ ID NO 17135
(R,K){20}
51: LHAIV KKKKKKKKKKKKKKKKKK KKKK
(K){20}
52: LHAIV KKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: HAVK KKKKKKKKKKKKKKKKKK EKKK
(K){20}
AAO03273 ck: 2981 len: 88 1 Aao03273 Human polypeptide SEQ ID NO 17165
(R,K){20}
43: IXYL KKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: XYLK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: YLKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: LKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
47: LKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
51: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}

52: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: KKKK KKKKKKKKKKKKKKKKKK KKKK
(K){20}
55: KKKK KKKKKKKKKKKKKKKKKK KKP
(K){20}
56: KKKK KKKKKKKKKKKKKKKKK KPPK
(K){20}
57: KKKK KKKKKKKKKKKKKKKKK PTPK
(K){20}
AAO03277 ck: 2981 len: 89 1 Aao03277 Human polypeptide SEQ ID NO 17
(R,K){20}
1: KKKKKKKKKKKKKKKKKK KKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
4: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
5: KKK KKKKKKKKKKKKKKKKK KKKK
(R,K){20}
AAO03278 ck: 6777 len: 74 1 Aao03278 Human polypeptide SEQ ID NO 17
(R,K){20}
47: RAVK KKKKKKKKKKKKKKKKK KKKK
(K){20}
48: AVAK KKKKKKKKKKKKKKKKK KKKK
(K){20}
49: VAKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
50: AKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
AAO03284 ck: 9650 len: 115 1 Aao03284 Human polypeptide SEQ ID NO 17
(R,K){20}
22: KALV KKKKKKKKKKKKKKKKK KKKK
(K){20}
23: ALVK KKKKKKKKKKKKKKKKK KKKK
(K){20}
24: LVSK KKKKKKKKKKKKKKKKK KKKK
(K){20}
25: VSKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
26: SKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
27: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}

```
(K){20}
28: KKKK KKKKKKKKKKKKKKKKKKK RGAL
(R,K){20}
29: KKKK KKKKKKKKKKKKKKKKKR GGALK

AA0328 ck: 1109 len: 117 1 Aa0328 Human polypeptide SEQ ID NO 17180.
(R,K){20}
31: SCSYP KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
32: CSYPR KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
33: SYPRK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
34: YPRKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
35: PRKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKR KKKK
(R,K){20}
47: KKKKK KKKKKKKKKKKKKKKRRR KKKK
(R,K){20}
48: KKKKK KKKKKKKKKKKKKKKRRR KKKK
(R,K){20}
49: KKKKK KKKKKKKKKKKKKKKRRR KKKK
(R,K){20}
50: KKKKK KKKKKKKKKKKKKKKRKK KKKK
(R,K){20}
51: KKKKK KKKKKKKKKKKKKRKKK KKTGG
```

```
(R,K){20}
52: KKKKK KKKKKKKKKKKKKKKKKK KTXGG
(R,K){20}
53: KKKKK KKKKKKKKKKKKKKKKKK TXGGG

AA03375 ck: 693 len: 54 1 Aa03375 Human polypeptide SEQ ID NO 172
(R,K){20}
25: TXLEB KKKKKKKKKKKKKKKKKK KKKAP
(K){20}
26: XLEBK KKKKKKKKKKKKKKKKKK KRAPG
(K){20}
27: LEEKK KKKKKKKKKKKKKKKKKK KAPGG
(K){20}
28: EEEKK KKKKKKKKKKKKKKKKKK APGGG

AA03459 ck: 5217 len: 58 1 Aa03459 Human polypeptide SEQ ID NO 173
(R,K){20}
28: XMCVF KKKKKKKKKKKKKKKKKK KDPPK
(K){20}
29: MCVFR KKKKKKKKKKKKKKKKKK DPEKK
(K){20}

AA03475 ck: 315 len: 138 1 Aa03475 Human polypeptide SEQ ID NO 173
(R,K){20}
44: LXLTL KKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: LXLTK KKKKKKKKKKKKKKKKKK KKKKG
(K){20}
46: XLTKK KKKKKKKKKKKKKKKKKK KKKGG
(K){20}
47: LTKKK KKKKKKKKKKKKKKKKKK KKGGP
(K){20}
48: TKKKK KKKKKKKKKKKKKKKKKK KGGBR
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKK GGPBG

AA03615 ck: 7613 len: 128 1 Aa03615 Human polypeptide SEQ ID NO 175
(R,K){20}
28: PLPRX KKKKKKKKKKKKKKKKKK KKTAA
(K){20}
29: LPXXK KKKKKKKKKKKKKKKKK KTAAG
(K){20}
30: PRXXK KKKKKKKKKKKKKKKKK KTAAG

AA03623 ck: 3884 len: 116 1 Aa03623 Human polypeptide SEQ ID NO 175
(R,K){20}
(K){20}
```

```

77: TPSPA KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
78: PSRAK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
79: SRAKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
80: RAKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
81: AKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
82: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}

AA003700 ck: 1279 len: 43 ! Aa03700 Human polypeptide SEQ ID NO.17592
      (R,K){20}
      (K){20}
19: YSQRL KKKKKKKKKKKKKKKKKKKKK GGGPF

AA003703 ck: 7016 len: 113 ! Aa03703 Human polypeptide SEQ ID NO.17595
      (R,R){20}
      (K){20}
58: IMDAE KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
59: WDAEK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
60: DAEKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
61: AEKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
62: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK SPGGA
      (K){20}

AA003766 ck: 8808 len: 91 ! Aa03766 Human polypeptide SEQ ID NO.17658

```

```

1      (R,K){20}
      (K){20}
49: TTAC KKKKKKKKKKKKKKKKKKK KKKK
50: TTAC KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
51: TACK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
52: ACKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
53: CKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
54: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
55: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
56: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
57: KKKK KKKKKKKKKKKKKKKKKKK GGGG
AA03841  ck: 8734  len: 100  ! Aa03841 Human polypeptide SEQ ID NO 17.
      (R,K){20}
      (K){20}
36: KOHP KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
37: QHPK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
38: HYPK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
39: YPKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
40: PKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
41: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
42: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
43: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
44: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (R,K){20}
45: KKKK KKKKKKKKKKKKKKKKKKK TQKK
AA03906  ck: 4312  len: 100  ! Aa03906 Human polypeptide SEQ ID NO 17.
      (R,K){20}
      (K){20}
34: NKON KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
35: KQNK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

```

```

36: ONQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
37: NOKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
38: OKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO03967 ck: 7917 len: 53 : Aao03967 Human polypeptide SEQ ID NO 17859
    (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: LKSKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KSKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KSKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: SKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO04512 ck: 4277 len: 62 : Aao04512 Human polypeptide SEQ ID NO 18404
    (R,K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

```

```

1
AAO04619 ck: 8654 len: 39 : Aao04619 Human polypeptide SEQ ID NO 185
    (R,K){20}
13: LIPEL KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: ITPEL KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
15: TPELK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
16: FLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AAO04644 ck: 2038 len: 70 : Aao04644 Human polypeptide SEQ ID NO 185
    (R,K){20}
23: AKPPT KKKKKKKKKKKKKKKKKKKKK KKKKK
    (R,K){20}

AAO04645 ck: 7825 len: 118 : Aao04645 Human polypeptide SEQ ID NO 185
    (R,K){20}
46: IKSPK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (R,K){20}
47: KSPFK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
48: SPFKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
49: FLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
50: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

```

58: KKKKK KKKKKKKKKKKKKKKKK KKKKP
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKK KKKPR
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKK KKKPRG
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKK KPRGG
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKK PRGGG
(K){20}

AA004647 ck: 4805 len: 58 1 Aa004647 Human polypeptide SEQ ID NO 18539.
(R,K){20}
28: KPTRP KKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: PTRPK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: TRPK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: RPKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: PKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKK S
(K){20}

AA004674 ck: 2036 len: 33 1 Aa004674 Human polypeptide SEQ ID NO 18566.
(R,K){20}
9: WCYIT KKKKKKKKKKKKKKKKK KKAAG
(K){20}
10: CYITK KKKKKKKKKKKKKKKKK KAGG
(K){20}
11: YITPK KKKKKKKKKKKKKKKKK ARG
(K){20}

AA004679 ck: 7190 len: 31 1 Aa004679 Human polypeptide SEQ ID NO 18571.
(R,K){20}
8: CWFTQ KKKKKKKKKKKKKKKKK KDGG
(K){20}
9: WFTQK KKKKKKKKKKKKKKKKK DRG
(K){20}

AA004682 ck: 5605 len: 60 1 Aa004682 Human polypeptide SEQ ID NO 185:
(R,K){20}
2: L KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: LK KKKKKKKKKKKKKKKKK KKQKD
(K){20}
4: LKK KKKKKKKKKKKKKKKKK KQKDL
(K){20}
5: LKKK KKKKKKKKKKKKKKKKK QKDLK
(K){20}

AA004690 ck: 7157 len: 81 1 Aa004690 Human polypeptide SEQ ID NO 18:
(R,K){20}
1: KKKKKKKKKKKKKKKKK KKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKK KKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKK PKKKK
(K){20}

AA004715 ck: 6984 len: 35 1 Aa004715 Human polypeptide SEQ ID NO 18:
(R,K){20}
8: LGSKD KKKKKKKKKKKKKKKKK KKAKK
(K){20}
9: GSKDK KKKKKKKKKKKKKKKKK KAKKK
(K){20}
10: SKDKK KKKKKKKKKKKKKKKKK AKKKD
(K){20}

AA004743 ck: 9412 len: 54 1 Aa004743 Human polypeptide SEQ ID NO 18:
(R,K){20}
21: KINKL KKKKKKKKKKKKKKKKK KRAAA
(K){20}
22: INKLK KKKKKKKKKKKKKKKKK RAAAR
(K){20}
23: NKLKK KKKKKKKKKKKKKKKKK AAARD
(R,K){20}

AA004747 ck: 8399 len: 39 1 Aa004747 Human polypeptide SEQ ID NO 18:

1

(R,K){20}

(K){20}

13: RTGFV KKKKKKKKKKKKKKKKKKK KRRGG

(K){20}

14: TGFVK KKKKKKKKKKKKKKKKKKK KRRGG

(K){20}

15: GFVKK KKKKKKKKKKKKKKKKKKK RGGGF

(R,K){20}

16: FVKKK KKKKKKKKKKKKKKKKKR GGGF

AA004752 ck: 1021 len: 71 1 Aa004752 Human polypeptide SEQ ID NO 18644.

1

(R,K){20}

(K){20}

19: QEGFL KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: EOGFL KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: OGLKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: GLKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(R){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKGG

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KGGGL

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKK GGGLL

AA004755 ck: 5521 len: 59 1 Aa004755 Human polypeptide SEQ ID NO 18647.

1

(R,K){20}

(K){20}

31: INSLE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: NSLEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

33: SLEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

34: LEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

35: EKKKK KKKKKKKKKKKKKKKKKKK KKKKT

(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKT

(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKK KKT

(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKK KT

1

(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKK T

AA004756 ck: 6952 len: 26 1 Aa004756 Human polypeptide SEQ ID NO 186

(R,K){20}

(K){20}

4: FFY KKKKKKKKKKKKKKKKKKK SSS

AA004758 ck: 7836 len: 115 1 Aa004758 Human polypeptide SEQ ID NO 186

1

(R,K){20}

(K){20}

7: PFYQL KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

8: FYQLK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

9: YQLKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

10: QLKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

11: LKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

27: KKKK (K){20} KKKK
28: KKKK (K){20} KKKK
29: KKKK (K){20} KKKK
30: KKKK (K){20} KKKK
31: KKKK (K){20} KKKK
32: KKKK (K){20} KKKK
33: KKKK (K){20} KKKK
34: KKKK (K){20} KKKK
35: KKKK (K){20} KKKK
36: KKKK (K){20} KKKK
37: KKKK (K){20} KKKK
38: KKKK (K){20} KKKK
39: KKKK (K){20} KKKK
40: KKKK (K){20} KKKK
41: KKKK (K){20} KKKK
42: KKKK (K){20} KKKK
43: KKKK (K){20} KKKK
44: KKKK (K){20} KKKK
45: KKKK (K){20} KKKK
46: KKKK (K){20} KKKK
47: KKKK (K){20} KKKK
48: KKKK (K){20} KKKK
49: KKKK (K){20} KKKK
50: KKKK (K){20} KKKK

1
AA04764 ck: 2035 len: 54 1 Aa04764 Human polypeptide SEQ ID NO 186
9: GDSSL (R,K){20} KKKK
10: DSSLK (K){20} KKKK
11: SSLK (K){20} KKKK
12: SLKK (K){20} KKKK
13: LKKK (K){20} KKKK
14: KKKK (K){20} KKKK
AA04802 ck: 1223 len: 105 1 Aa04802 Human polypeptide SEQ ID NO 186
34: FFSRQ (R,K){20} KKKK
AA04854 ck: 4121 len: 58 1 Aa04854 Human polypeptide SEQ ID NO 187
13: TPFRQ (R,K){20} KKKK
14: PFRK (K){20} KKKK
15: FRKK (K){20} KKKK
16: RAKK (K){20} KKKK
17: AKKK (K){20} KKKK
18: KKKK (K){20} KKKK
19: KKKK (K){20} KKKK
20: KKKK (K){20} KKKK
21: KKKK (K){20} KKKK
22: KKKK (K){20} KKKK
23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
25: KKKK (K){20} KKKK

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKRGG
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KRGGN
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK RGGNF
(K){20}
32: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKR GGNFK

AA004856 ck: 4861 len: 135 ! Aa004856 Human polypeptide SEQ ID NO 18748.

41: LTLTT (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: TLTTK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: LTTKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: TTKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: TKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKRGG
(K){20}
51: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKGGG
(K){20}
52: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KGGGG
(K){20}
53: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK GGGGP
(K){20}

AA004872 ck: 3224 len: 93 ! Aa004872 Human polypeptide SEQ ID NO 18764.

(R,K){20}
(K){20}
8: FLYKL KKKKKKKKKKKKKKKKKKK KKKRG
(K){20}
9: LYKLK KKKKKKKKKKKKKKKKKKK KKRGG
(K){20}

(K){20}
10: YKLKK KKKKKKKKKKKKKKKKKKK KRGGP
(K){20}
11: KLKKK KKKKKKKKKKKKKKKKKKK RGGPL
(R,K){20}
12: LKKKK KKKKKKKKKKKKKKKKKKK GGPLK
(R,K){20}

AA004873 ck: 7719 len: 66 ! Aa004873 Human polypeptide SEQ ID NO 187

(R,K){20}
(K){20}
21: SFLEK KKKKKKKKKKKKKKKKKKK KGGGP
(K){20}
22: FLIEK KKKKKKKKKKKKKKKKKKK GGGPL
(K){20}

AA004874 ck: 5753 len: 58 ! Aa004874 Human polypeptide SEQ ID NO 187

(R,K){20}
(K){20}
10: SEKLP KKKKKKKKKKKKKKKKKKK KKGKK
(K){20}
11: EKLPK KKKKKKKKKKKKKKKKKKK KGKLI
(K){20}
12: KLPPK KKKKKKKKKKKKKKKKKKK GKLIK
(K){20}

AA004881 ck: 8841 len: 115 ! Aa004881 Human polypeptide SEQ ID NO 187

(R,K){20}
(K){20}
24: ITPHP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: TPHPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: PPHPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: HPKKK KKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
28: PKKKK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKGGA
(K){20}
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KGAL
(K){20}
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK GGALK
(K){20}

AA004917 ck: 3999 len: 103 ! Aa004917 Human polypeptide SEQ ID NO 188

(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKKKK GGGPL
(K){20}

AA004928 ck: 23 len: 43 ! Aa004928 Human polypeptide SEQ ID NO 188

(R,K){20}
(K){20}
19: FKKEK KKKKKKKKKKKKKKKKKKK GGGGF
(K){20}

1
AA004969 ck: 4345 len: 57 ! Aa004969 Human polypeptide SEQ ID NO 18861
(R,K){20}

26: LMGTS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

27: MGTSK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

28: GTSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

29: TSRRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

30: SKRRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

33: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

34: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}

1
AA005000 ck: 5847 len: 38 ! Aa005000 Human polypeptide SEQ ID NO 18992
(R,K){20}

5: SGRS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

6: SGRSK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

7: GRSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

8: RSRRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: SKRRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}

13: KKKKK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KRRGR
(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK RGR
(R,K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK RGR
(R,K){20}

1
AA005081 ck: 2260 len: 40 ! Aa005081 Human polypeptide SPQ ID NO 18
(R,K){20}

9: SRASP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: RASPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: ASPKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

12: SPKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

13: PKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKS
(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKK KKKSS
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK KSSK
(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKK KSSK
(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKK SSK
(K){20}

1
AA005130 ck: 6597 len: 116 ! Aa005130 Human polypeptide SEQ ID NO 19
(R,K){20}

14: ETPFQ KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}

15: TPFQK KKKKKKKKKKKKKKKKKKK RCGGF
(K){20}

16: PFGKK KKKKKKKKKKKKKKKKKKK GGGFL
(R,K){20}

1
AA005191 ck: 4399 len: 85 ! Aa005191 Human polypeptide SEQ ID NO 19
(R,K){20}

30: GYIQL KKKKKKKKKKKKKKKKKKK RGAFF
(R,K){20}

1
AA005260 ck: 3618 len: 55 ! Aa005260 Human polypeptide SEQ ID NO 19
(R,K){20}

15: PHKHQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}

```
16: HHHQK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
17: KHQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
18: HQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
19: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGKG
    (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKK KGGGD
    (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKK GGSGD
    (K){20}

AA005369 ck: 1417 len: 70 1 Aa005369 Human polypeptide SEQ ID NO 19261.
    (R,K){20}
22: KYHET KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
23: YHETK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
24: HETKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
25: ETKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
26: TKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGE
    (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGKK
    (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KGKKK
    (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK GKKKK
    (R,K){20}

AA005372 ck: 9880 len: 43 1 Aa005372 Human polypeptide SEQ ID NO 19264.
    (R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
2: K KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
3: KK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
```

```
5: KKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
11: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AA005384 ck: 715 len: 23 1 Aa005384 Human polypeptide SEQ ID NO 192
    (R,K){20}
4: LFS KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}

AA005499 ck: 7813 len: 76 1 Aa005499 Human polypeptide SEQ ID NO 193
    (R,K){20}
7: VQPOQ KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
8: QPOQK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
9: POQKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
10: QQKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
11: QKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
    (K){20}
```

15: KKKKK (K){20} KKKKK
16: KKKKK (K){20} KKKKK
17: KKKKK (K){20} KKKKK
18: KKKKK (K){20} KKKKK
19: KKKKK (K){20} KKKKK
20: KKKKK (K){20} KKKKK
21: KKKKK (K){20} KKKKK
22: KKKKK (K){20} KKTAS
23: KKKKK (K){20} KKASP
24: KKKKK (K){20} KASPG
25: KKKKK (K){20} ASPGG
AAO05500 ck: 7443 len: 53 ! Aao05500 Human polypeptide SEQ ID NO 19392.
(R,K){20}
(K){20}
5: QPOS KKKKKKKKKKKKKKKKK KASQ
(K){20}
6: QPOS KKKKKKKKKKKKKKKKK KASQK
(K){20}
7: POSK KKKKKKKKKKKKKKKKK ASQKK
(K){20}
AAO05530 ck: 7354 len: 75 ! Aao05530 Human polypeptide SEQ ID NO 19422.
(R,K){20}
(K){20}
11: TTYFL KKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: TYFLK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: YFLK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: FLKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: LKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

18: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
39: KKKKK KKKKKKKKKKKKKKKKK KKKAS
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKK KKTASS
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKK KASSP
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKK ASSPP
(K){20}

100: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AA007505 ck: 3528 len: 72 1 Aa007505 Human polypeptide SEQ ID NO 21397.

(R,K){20}
(K){20}

14: HLVEA KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: LVEAK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: VEAKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: EAKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

AA007509 ck: 4998 len: 140 1 Aa007509 Human polypeptide SEQ ID NO 21401.

(R,K){20}

17: KVKKE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: VKVEK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KVEKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: XEKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

(K){20}

21: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

32: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

AA007594 ck: 7271 len: 93 1 Aa007594 Human polypeptide SEQ ID NO 21.

(R,K){20}

9: FLLGG KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

10: LLGGK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

11: LLGGK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

12: LGKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

13: GKKKK KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

AA007607 ck: 2434 len: 72 1 Aa007607 Human polypeptide SEQ ID NO 21.

(R,K){20}

21: KKKKE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

48: GLFKE KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

AA007610 ck: 5270 len: 74 1 Aa007610 Human polypeptide SEQ ID NO 21.

(R,K){20}

19: TELTI KKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

(K){20}

(K){20}

20: ELTIK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: LTIKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: TIKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: IKKKK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKGCG
(K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KGGGA
(K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK GG GAL

AAO07755 ck: 6069 len: 90 i Aao07755 Human polypeptide SEQ ID NO 21647.

(R,K){20}
35: KNOSS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
36: NQSSK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
37: QSSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
38: SSKKK KKKKKKKKKKKKKKKKKKK KKKKG
(K){20}
39: SKKKK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKGGP
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KGGPL
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK GGPLL

AAO07762 ck: 5115 len: 41 i Aao07762 Human polypeptide SEQ ID NO 21654.

(R,K){20}
18: IPSLX KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: PSLXX KKKKKKKKKKKKKKKKKKK KKK
(K){20}
20: SLKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
21: LKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
22: XKKKK KKKKKKKKKKKKKKKKKKK

AAO07775 ck: 9827 len: 90 i Aao07775 Human polypeptide SEQ ID NO 21667.

(R,K){20}
46: PHXPE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
47: HXPEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: XPEKK KKKKKKKKKKKKKKKKKKK KKKGG
(K){20}
49: PEKKK KKKKKKKKKKKKKKKKKKK KKGGA
(K){20}
50: EKKKK KKKKKKKKKKKKKKKKKKK KGGAF
(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK GGAFK

AAO07806 ck: 1959 len: 99 i Aao07806 Human polypeptide SEQ ID NO 216

(R,K){20}
41: KLHCI KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: LHCIK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: HCIKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: CIKKK KKKKKKKKKKKKKKKKKKK KKKKL
(K){20}
45: IKKKK KKKKKKKKKKKKKKKKKKK KKKLW
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKK KKLWG
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKK KKLWG
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKK LMGGG

AAO07872 ck: 8003 len: 135 i Aao07872 Human polypeptide SEQ ID NO 217

(R,K){20}
52: TKKKQ KKKKKKKKKKKKKKKKKKK GGGLL
(K){20}

AAO07874 ck: 8007 len: 111 i Aao07874 Human polypeptide SEQ ID NO 217

(R,K){20}
90: EEPST KKKKKKKKKKKKKKKKKKK KG
(K){20}
91: EPSIK KKKKKKKKKKKKKKKKKKK G
(K){20}

AAO07876 ck: 8130 len: 66 i Aao07876 Human polypeptide SEQ ID NO 217

(R,K){20}
17: IYSPF KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: YSPEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: SPEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
20: FEKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: EKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKK KKKKK

AA007976 ck: 6471 len: 97 ! Aa007976 Human polypeptide SEQ ID NO 21868.

(R,K){20}

43: FPIST KKKKKKKKKKKKKKKKK KKKKK
(K){20}

44: PISTK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

45: ISTKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

46: STKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

47: TKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

48: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

49: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

50: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

51: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

52: KKKKK KKKKKKKKKKKKKKKKK KKKAS
(K){20}

53: KKKKK KKKKKKKKKKKKKKKKK KKKAS
(K){20}

54: KKKKK KKKKKKKKKKKKKKKKK KASSG
(K){20}

55: KKKKK KKKKKKKKKKKKKKKKK ASSGG
(K){20}

AA007981 ck: 9551 len: 81 ! Aa007981 Human polypeptide SEQ ID NO 21873.

(R,K){20}

54: KPCIY KKKKKKKKKKKKKKKKK KKKKK
(K){20}

55: PCIYK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

56: CIYKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

57: IYKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

58: YKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKK KKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKK KK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKK K
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKK

AA008018 ck: 9768 len: 48 ! Aa008018 Human polypeptide SEQ ID NO 219

(R,K){20}

24: KISQV KKKKKKKKKKKKKKKKK KKKKK
(R,K){20}

25: ISOVR KKKKKKKKKKKKKKKKK KKKKK
(K){20}

26: SQVRK KKKKKKKKKKKKKKKKK KKK
(K){20}

27: QVRKK KKKKKKKKKKKKKKKKK KK
(K){20}

28: VRKKK KKKKKKKKKKKKKKKKK K
(K){20}

29: RKKKK KKKKKKKKKKKKKKKKK
(K){20}

AA008121 ck: 5381 len: 86 ! Aa008121 Human polypeptide SEQ ID NO 220

(R,K){20}

20: HFGLL KKKKKKKKKKKKKKKKK GGCAF
(K){20}

AA008232 ck: 782 len: 104 ! Aa008232 Human polypeptide SEQ ID NO 221

(R,K){20}

49: ISQON KKKKKKKKKKKKKKKKK KDSGG
(K){20}

50: SQCNK KKKKKKKKKKKKKKKKK DGGGA
(K){20}

AA008450 ck: 6228 len: 31 ! Aa008450 Human polypeptide SEQ ID NO 223

(R,K){20}

5: NNSA KKKKKKKKKKKKKKKKK KKGGA
(K){20}

6: NNSAK KKKKKKKKKKKKKKKKK KGGAA
(K){20}

7: NSAKK KKKKKKKKKKKKKKKKK GGAAF
(K){20}

AA008458 ck: 1813 len: 59 ! Aa008458 Human polypeptide SEQ ID NO 223

(R,K){20}

8: LGWE KKKKKKKKKKKKKKKKK KKKKK
(K){20}

(K){20}
9: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: GWEK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: WEKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: EKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
AA008460 ck: 5334 len: 62 1 Aa008460 Human polypeptide SEQ ID NO 22352.
(R,K){20}
22: KEFT KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
23: EFST KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
24: FSTK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
25: STKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
26: TKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
27: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
28: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
30: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
31: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
32: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
33: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
34: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
AA008466 ck: 8728 len: 107 1 Aa008466 Human polypeptide SEQ ID NO 22358.
(R,K){20}
57: FWGV KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
58: WGVK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
59: GGVK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
60: GVKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}
61: VKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
62: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
63: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
64: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
65: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
AA008580 ck: 5164 len: 62 1 Aa008580 Human polypeptide SEQ ID NO 224
(R,K){20}
30: PILL KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AA008591 ck: 900 len: 46 1 Aa008591 Human polypeptide SEQ ID NO 224
(R,K){20}
14: LLCS KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AA008609 ck: 5151 len: 74 1 Aa008609 Human polypeptide SEQ ID NO 225
(R,K){20}
18: KCVL KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
19: CVLK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AA008612 ck: 557 len: 46 1 Aa008612 Human polypeptide SEQ ID NO 225
(R,K){20}
20: TFCM KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
21: FCIM KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
22: CIMK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
23: IMKK KKKKKKKKKKKKKKKKKKKKKKKKK
AA008616 ck: 4555 len: 41 1 Aa008616 Human polypeptide SEQ ID NO 225
(R,K){20}
6: HCALP KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: CALP KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: ALPK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: LPKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

10: PKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK
(K){20}

AA008623 ck: 9243 len: 119 1 Aa008623 Human polypeptide SEQ ID NO 22515.
(R,K){20}
(R,K){20}

21: DSKQZ KKKKKKKKKKKKKKKKKKKR GGAFK
(R,K){20}

AA008624 ck: 5590 len: 65 1 Aa008624 Human polypeptide SEQ ID NO 22516.
(R,K){20}
(K){20}

11: EGNMW KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

12: GNNWK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

13: NNWKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

14: NNWKK KKKKKKKKKKKKKKKKKKK KKWGG
(K){20}

15: WKKKK KKKKKKKKKKKKKKKKKKK KWWGA
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKK WGGAL
(K){20}

AA008631 ck: 6430 len: 135 1 Aa008631 Human polypeptide SEQ ID NO 22523.
(R,K){20}
(K){20}

30: TPSRA KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

31: PSRAK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

32: SRAKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

33: RAKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

34: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

39: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

40: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKK KRRGG
(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKK RGGGP
(R,K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKR GGGP
(R,K){20}

AA008644 ck: 3202 len: 80 1 Aa008644 Human polypeptide SEQ ID NO 225
(R,K){20}
(K){20}

2: E KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

3: EK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

4: EKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

5: EKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

6: EKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

11: KKKKK (K){20} KKKKK
12: KKKKK (K){20} KKKKK
13: KKKKK (K){20} KKKKK
14: KKKKK (K){20} KKKKK
15: KKKKK (K){20} KKKKK
16: KKKKK (K){20} KKKGG
17: KKKKK (K){20} KKKGG
18: KKKKK (K){20} KGGGG
19: KKKKK (K){20} GGGGP
AAO08653 ck: 6448 len: 63 1 Aao08653 Human polypeptide SEQ ID NO 22545

12: PIRLT (R,K){20} KKKKK
13: IKLTG (K){20} KKKKK
14: KLTGK (K){20} KKKKK
15: LTGKK (K){20} KKKKK
16: TKGKK (K){20} KKKKK
17: KKKKK (K){20} KKKGA
18: KKKKK (K){20} KKGAP
19: KKKKK (K){20} KGAPL
20: KKKKK (R,K){20} GADLK

1 AAO08686 ck: 6936 len: 81 1 Aao08686 Human polypeptide SEQ ID NO 22578
(R,K){20}
(K){20}
23: HPVNH KKKKK KKKKK
24: PVNHK (K){20} KKKKK
25: VNHKK (K){20} KKKKK
(K){20}

26: NHKKK KKKKK KKKKK
27: HKKKK (K){20} KKKKK
28: KKKKK (K){20} KKKKK
29: KKKKK (K){20} KKKKK
30: KKKKK (K){20} KKKKK
31: KKKKK (K){20} KKKKK
32: KKKKK (K){20} KKKKK
33: KKKKK (K){20} KKKKE
34: KKKKK (K){20} KKKEG
35: KKKKK (K){20} KKEGA
36: KKKKK (K){20} KEGAL
37: KKKKK (K){20} EGALG
AAO08701 ck: 588 len: 46 1 Aao08701 Human polypeptide SEQ ID NO 225
(R,K){20}
(K){20}
7: INANS KKKKK KKKKK
8: NANSK (K){20} KKKKK
9: ANSKK (K){20} KKKKG
10: NSKKK (K){20} KKKGG
11: SKKKK (K){20} KKKGG
12: KKKKK (K){20} KKGGL
13: KKKKK (K){20} GGGLL
AAO08707 ck: 9359 len: 48 1 Aao08707 Human polypeptide SEQ ID NO 225
(R,K){20}
(K){20}
13: TAVTQ KKKKK KKKKK
14: AVTQK (K){20} KGGGV
15: VTQKK (K){20} GGGLL

1
AAO08708 ck: 9654 len: 32 ! Aao08708 Human polypeptide SEQ ID NO 22600.
(R,K){20}
(K){20}
8: SMTEP KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: MTEPK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: TEPKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
12: PPKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08717 ck: 4688 len: 99 ! Aao08717 Human polypeptide SEQ ID NO 22609.
(R,K){20}
(K){20}
80: SPILLI KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
1
AAO08772 ck: 1431 len: 105 ! Aao08772 Human polypeptide SEQ ID NO 22664.
(R,K){20}
(K){20}
23: IMNYL KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
24: MNYLK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
25: NYLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
26: YLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
27: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08774 ck: 2529 len: 71 ! Aao08774 Human polypeptide SEQ ID NO 22666.
(R,K){20}
(K){20}
16: RTIKL KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
17: TIKLK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
18: IKLKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
19: KLKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1
20: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08817 ck: 529 len: 28 ! Aao08817 Human polypeptide SEQ ID NO 227.
(R,K){20}
(K){20}
3: AX KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08820 ck: 1888 len: 84 ! Aao08820 Human polypeptide SEQ ID NO 227.
(R,K){20}
(K){20}
28: CMTPS KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: MTFKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}
AAO08841 ck: 3529 len: 74 ! Aao08841 Human polypeptide SEQ ID NO 227.
(R,K){20}
(K){20}
3: PQ KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: POK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08857 ck: 2816 len: 49 ! Aao08857 Human polypeptide SEQ ID NO 227.
(R,K){20}
(K){20}
25: PPPTS KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
26: PPTSK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
27: PLSKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
28: TSKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
29: SKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
AAO08913 ck: 3132 len: 52 ! Aao08913 Human polypeptide SEQ ID NO 228.
(R,K){20}
(K){20}
9: IIKSF KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: IKSPK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KSFKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

12: SFRKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: FKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO08914 ck: 4641 len: 25 ! Aao08914 Human polypeptide SEQ ID NO 22806
(R,K){20}
(R,K){20}
6: ILMPX KKKKKKKKKKKKKKKKKKK KKKKK
AAO08943 ck: 5770 len: 75 ! Aao08943 Human polypeptide SEQ ID NO 22835
(R,K){20}
(K){20}

8: SEMAA KKKKKKKKKKKKKKKKKKK KKKKK
9: EWAAR (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
10: WAAKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: AAARR (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO08994 ck: 2695 len: 29 ! Aao08994 Human polypeptide SEQ ID NO 228
(R,K){20}
(K){20}
9: NLPKH KKKKKKKKKKKKKKKKKKK KKKKK
10: LPSHK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO08995 ck: 6270 len: 26 ! Aao08995 Human polypeptide SEQ ID NO 228
(R,K){20}
(K){20}
6: MAAPP KKKKKKKKKKKKKKKKKKK KKKKK
AAO09001 ck: 6400 len: 26 ! Aao09001 Human polypeptide SEQ ID NO 228
(R,K){20}
(K){20}
5: LTSS KKKKKKKKKKKKKKKKKKK KKKKK
6: LTSSK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
7: TSSKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO09016 ck: 4378 len: 119 ! Aao09016 Human polypeptide SEQ ID NO 229
(R,K){20}
(R,K){20}
10: EEEEE RRRRRRRRRRRRRRRRRR ILRKK
AAO09057 ck: 3712 len: 58 ! Aao09057 Human polypeptide SEQ ID NO 229
(R,K){20}
(K){20}
5: MILN KKKKKKKKKKKKKKKKKKK KKKKK
6: MILNK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
7: ILNKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
8: LNKRR (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
9: NKKRR (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
10: KKKRR (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
AAO09066 ck: 2645 len: 29 ! Aao09066 Human polypeptide SEQ ID NO 229

```
1
(R,K){20}
4: AQQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
5: AQQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
6: AQQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
7: QQQ KKKKKKKKKKKKKKKKKKK KKK
(K){20}
8: QKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
9: KKK KKKKKKKKKKKKKKKKKKK K
(K){20}
10: KKK KKKKKKKKKKKKKKKKKKK
(K){20}

AAO09072 ck: 8432 len: 42 1 Aao09072 Human polypeptide SEQ ID NO 22964.
(R,K){20}
11: LIPTL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: IFTLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: FTLLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: TLKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAO09077 ck: 9524 len: 87 1 Aao09077 Human polypeptide SEQ ID NO 22969.
(R,K){20}
26: RFLTL KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: FLTLK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: LLTK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: LTKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: TKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAO09162 ck: 3841 len: 100 1 Aao09162 Human polypeptide SEQ ID NO 23054.
(R,K){20}
11: KTLFQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: TLFQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
13: LFQK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
14: FQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

```
1
(K){20}
15: QKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
16: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
17: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
18: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
19: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
23: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAO09258 ck: 7988 len: 42 1 Aao09258 Human polypeptide SEQ ID NO 231.
(R,K){20}
7: LFCP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
8: LFCP KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
9: FCPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
10: CFPK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
11: PKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
12: KKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAO09269 ck: 7553 len: 106 1 Aao09269 Human polypeptide SEQ ID NO 232.
(R,K){20}
39: NTGML KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

1 AAC09457 ck: 9658 len: 32 ! Aao09457 Human polypeptide SEQ ID NO 23349.
(R,K){20}
(K){20}
9: ALVPQ KKKKKKKKKKKKKKKKKKK NIKI

1 AAC09819 ck: 4709 len: 80 ! Aao09819 Human polypeptide SEQ ID NO 23711.
(R,K){20}
(K){20}
54: CEFVX KKKKKKKKKKKKKKKKKKK EKGGS

1 AA010447 ck: 6342 len: 50 ! Aao10447 Human polypeptide SEQ ID NO 24339.
(R,K){20}
(K){20}
25: RFLKL KKKKKKKKKKKKKKKKKKK RKKKK
(R,K){20}
26: FLKLK KKKKKKKKKKKKKKKKKKK KKKKI
(R,K){20}
27: LKLKK KKKKKKKKKKKKKKKKKKK KKKI
(R,K){20}
28: KLKKK KKKKKKKKKKKKKKKKKKK KKI
(R,K){20}
29: LKKKK KKKKKKKKKKKKKKKKKKK KI
(R,K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK I

1 AA010451 ck: 3955 len: 82 ! Aao10451 Human polypeptide SEQ ID NO 24343.
(R,K){20}
(K){20}
59: SRASP KKKKKKKKKKKKKKKKKKK ARGG

1 AA010467 ck: 7542 len: 116 ! Aao10467 Human polypeptide SEQ ID NO 24359.
(R,K){20}
(K){20}
57: KCEFM KKKKKKKKKKKKKKKKKKK KIGGG
(K){20}
58: CEFMK KKKKKKKKKKKKKKKKKKK IGGGA

1 AA010564 ck: 9156 len: 98 ! Aao10564 Human polypeptide SEQ ID NO 24456.
(R,K){20}
(K){20}
18: NLTLT KKKKKKKKKKKKKKKKKKK KKEMP
(R,K){20}
19: LTLTK KKKKKKKKKKKKKKKKKKK KEMPV
(K){20}
20: LTLKK KKKKKKKKKKKKKKKKKKK EMPVK

1 AA010608 ck: 9106 len: 67 ! Aao10608 Human polypeptide SEQ ID NO 24500.
(R,K){20}
(K){20}
42: CRUSE KKKKKKKKKKKKKKKKKKK KKKKD

1 43: RLSEK KKKKKKKKKKKKKKKKKKK KKKDS
(K){20}
44: LSEKK KKKKKKKKKKKKKKKKKKK KKDS
(K){20}
45: SEKKK KKKKKKKKKKKKKKKKKKK KDS
(K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKK DS

1 AA010638 ck: 621 len: 56 ! Aao10638 Human polypeptide SEQ ID NO 245
(R,K){20}
(K){20}
33: CEPQP KKKKKKKKKKKKKKKKKKK KARG
(K){20}
34: EPQPK KKKKKKKKKKKKKKKKKKK ARG

1 AA010786 ck: 7349 len: 65 ! Aao10786 Human polypeptide SEQ ID NO 246
(R,K){20}
(K){20}
24: CPKFS KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: PKFSK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: XFSKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
27: FSKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
28: SKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKV
(K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKK KKKVF
(K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKK KKVFE
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKK KVFEE
(K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKK VFEEE

1 AA010852 ck: 5147 len: 41 ! Aao10852 Human polypeptide SEQ ID NO 247
(R,K){20}
(K){20}
20: FSRKM KKKKKKKKKKKKKKKKKKK KK
(K){20}
21: SRMKK KKKKKKKKKKKKKKKKKKK K

22: RYMKR (K){20}
XXXXXXXXXXXXXXXXXXXX

AA010853 ck: 444 len: 98 ! Aa010853 Human polypeptide SEQ ID NO 24745.

14: LRAHL (R,K){20}
XXXXXXXXXXXXXXXXXXXX

15: RAHLK (K){20}
XXXXXXXXXXXXXXXXXXXX

16: AHLKK (K){20}
XXXXXXXXXXXXXXXXXXXX

17: HLKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

18: LKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

19: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

20: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

21: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

22: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

23: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

24: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

25: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

26: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

27: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

28: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

29: KKKKK (R,K){20}
XXXXXXXXXXXXXXXXXXXX

AA010859 ck: 2690 len: 70 ! Aa010859 Human polypeptide SEQ ID NO 24751.

44: XGITE (R,K){20}
XXXXXXXXXXXXXXXXXXXX

45: GITEK (K){20}
XXXXXXXXXXXXXXXXXXXX

AA010933 ck: 849 len: 69 ! Aa010933 Human polypeptide SEQ ID NO 24825.

33: FINTE (R,K){20}
XXXXXXXXXXXXXXXXXXXX

(K){20}

34: INTEK KKKKKKKKKKKKKKKK

35: NTEKK (K){20}
XXXXXXXXXXXXXXXXXXXX

36: TEKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

37: EKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

38: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

39: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

40: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

41: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

42: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

43: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

44: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

45: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

AA010973 ck: 824 len: 51 ! Aa010973 Human polypeptide SEQ ID NO 248

16: SYFYM (R,K){20}
XXXXXXXXXXXXXXXXXXXX

17: YFYMK (R,K){20}
XXXXXXXXXXXXXXXXXXXX

AA010983 ck: 694 len: 51 ! Aa010983 Human polypeptide SEQ ID NO 248

24: HPILX (R,K){20}
XXXXXXXXXXXXXXXXXXXX

25: PILXK (K){20}
XXXXXXXXXXXXXXXXXXXX

26: ILXKK (K){20}
XXXXXXXXXXXXXXXXXXXX

27: LXXKK (K){20}
XXXXXXXXXXXXXXXXXXXX

28: XXXKK (K){20}
XXXXXXXXXXXXXXXXXXXX

29: KKKKK (K){20}
XXXXXXXXXXXXXXXXXXXX

AA010995 ck: 2382 len: 122 ! Aa010995 Human polypeptide SEQ ID NO 248

22: PCKGS (R,K){20}
XXXXXXXXXXXXXXXXXXXX

(K){20}

(K){20}
23: CKGSK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
24: KGSKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
25: GSKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
26: SKKKK KKKKKKKKKKKKKKKKK KKGCG
(K){20}
27: KKKKK KKKKKKKKKKKKKKKKK KGGGG
(K){20}
28: KKKKK KKKKKKKKKKKKKKKKK GGGGF

AA01097 ck: 5315 len: 58 1 Aa01097 Human polypeptide SEQ ID NO 24899.
(R,K){20}
29: IYIIQ KKKKKKKKKKKKKKKKK KKKKK
(K){20}
30: XIIOK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
31: IIOKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
32: IOKKK KKKKKKKKKKKKKKKKK KRRGG
(K){20}
33: QKKKK KKKKKKKKKKKKKKKKK KRRGG
(K){20}
34: KKKKK KKKKKKKKKKKKKKKKK RGGPP
(R,K){20}
35: KKKKK KKKKKKKKKKKKKKKKK GGPP

AA01102 ck: 8638 len: 82 1 Aa01102 Human polypeptide SEQ ID NO 24894.
(R,K){20}
5: FKRS KKKKKKKKKKKKKKKKKR GGALL
(R,K){20}
38: XFKRS KKKKKKKKKKKKKKKKKR GGALL

AA01103 ck: 3780 len: 49 1 Aa01103 Human polypeptide SEQ ID NO 24925.
(R,K){20}
10: QKKFI RKKKKKKKKKKKKKKKK KLSKK
(R,K){20}
11: XKFIR KKKKKKKKKKKKKKKKK LSKFI
(K){20}

AA01104 ck: 7868 len: 85 1 Aa01104 Human polypeptide SEQ ID NO 24940.
(R,K){20}
53: IISDP KKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: ISDPK KKKKKKKKKKKKKKKKK KKKKK

1

(K){20}
55: SDPKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: DPKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: PKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}

AA01107 ck: 4186 len: 100 1 Aa01107 Human polypeptide SEQ ID NO 249
(R,K){20}
39: PPKPD KKKKKKKKKKKKKKKKK KKKKK
(K){20}
40: PKPKD KKKKKKKKKKKKKKKKK KKKKK
(K){20}
41: KPDKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
42: PDKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
43: DKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
44: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: KKKKK KKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKK KRRGG

50: KKKK (K){20} KRGG
51: KKKK (K){20} RGGG
52: KKKK (R,K){20} GGGG
AA01124 ck: 5121 len: 61 ! Aa01124 Human polypeptide SEQ ID NO 25016.

1
16: KYSFL (R,K){20} KKKK
17: YSFLK (K){20} KKKK
18: SFLK (K){20} KKKK
19: FLKK (K){20} KKKK
20: LKKK (K){20} KKKK
21: KKKK (K){20} KKKK
22: KKKK (K){20} KKKK
23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
25: KKKK (K){20} KKKK
26: KKKK (K){20} KKKK
27: KKKK (K){20} KKKK
28: KKKK (K){20} KKKK
29: KKKK (K){20} KKKK
30: KKKK (K){20} KKKK
31: KKKK (K){20} KKKK
32: KKKK (K){20} KKKK
33: KKKK (K){20} KKKK
34: KKKK (K){20} KKKK
35: KKKK (K){20} KKKK

36: KKKK (K){20} KKKK
37: KKKK (K){20} KKKK
38: KKKK (K){20} KKKK
39: KKKK (K){20} KKKK
40: KKKK (K){20} KK
41: KKKK (K){20} K
42: KKKK (K){20} KKKK

1
AA01139 ck: 8807 len: 68 ! Aa01139 Human polypeptide SEQ ID NO 250
22: VCFEP (R,K){20} KKEGG
23: CREPK (K){20} MEGG

1
AA01165 ck: 7203 len: 62 ! Aa01165 Human polypeptide SEQ ID NO 250
18: DASV (R,K){20} KKKK
19: ASVVK (K){20} KKKK
20: SVVK (K){20} KKKK
21: MVKK (K){20} KKKK
22: VKKK (K){20} KKKK
23: KKKK (K){20} KKKK
24: KKKK (K){20} KKKK
25: KKKK (K){20} KASS
26: KKKK (K){20} ASSQ

1
AA01209 ck: 6562 len: 96 ! Aa01209 Human polypeptide SEQ ID NO 251
63: LLAY (R,K){20} KKKK
64: LLAVK (K){20} KKKK

40: SARKK (K){20}
KKKKKKKKKKKKKKKKKK GGGGV

AA011327 ck: 3052 len: 102 ! Aa011327 Human polypeptide SEQ ID NO 25219.

(R,K){20}
20: LSKEL (K){20}
KKKKKKKKKKKKKKKKKK KKKLG

(K){20}
21: SKELK (K){20}
KKKKKKKKKKKKKKKKKK KKLGD

(K){20}
22: KELKK (K){20}
KKKKKKKKKKKKKKKKKK KLGDE

(K){20}
23: ELKKK (K){20}
KKKKKKKKKKKKKKKKKK LGDEE

AA011342 ck: 7189 len: 106 ! Aa011342 Human polypeptide SEQ ID NO 25234.

(R,K){20}
61: FFKTX (K){20}
KKKKKKKKKKKKKKKKKK GGPKL

AA011346 ck: 7311 len: 85 ! Aa011346 Human polypeptide SEQ ID NO 25238.

(R,K){20}
62: ANHWE (K){20}
KKKKKKKKKKKKKKKKKK SGGG

AA011352 ck: 1342 len: 51 ! Aa011352 Human polypeptide SEQ ID NO 25244.

(R,K){20}
25: SIFKX (K){20}
KKKKKKKKKKKKKKKKKK KNPPS

(K){20}
26: IFKKK (K){20}
KKKKKKKKKKKKKKKKKK NPPSL

AA011705 ck: 7660 len: 101 ! Aa011705 Human polypeptide SEQ ID NO 25597.

(R,K){20}
17: WAPLQ (K){20}
KKKKKKKKKKKKKKKKKK KKKGG

(K){20}
18: APLQK (K){20}
KKKKKKKKKKKKKKKKKK KKGGA

(K){20}
19: PLQKK (K){20}
KKKKKKKKKKKKKKKKKK KKGAP

(K){20}
20: LQKKK (K){20}
KKKKKKKKKKKKKKKKKK GGAPL

AA011799 ck: 8286 len: 39 ! Aa011799 Human polypeptide SEQ ID NO 25691.

(R,K){20}
7: NPPVS (R,K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
8: PPSVR (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
9: PPSRK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
10: VSRKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
11: SRKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
12: RKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
13: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKGG

(K){20}
14: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKGCG

(K){20}
15: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KGGGL

(K){20}
16: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK GGGL

AA011820 ck: 8725 len: 42 ! Aa011820 Human polypeptide SEQ ID NO 257

(R,K){20}
17: NFHAL (K){20}
KKKKKKKKKKKKKKKKKK SSSLR

AA011828 ck: 4807 len: 41 ! Aa011828 Human polypeptide SEQ ID NO 257

(R,K){20}
1: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
2: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
3: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
4: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKKK

(K){20}
5: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKRG

(K){20}
6: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKRGG

(K){20}
7: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KRGGG

(K){20}
8: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK RGGGF

(R,K){20}
9: KKKKK (R,K){20}
KKKKKKKKKKKKKKKKKK GGGFK

AA011831 ck: 1761 len: 29 ! Aa011831 Human polypeptide SEQ ID NO 257

(R,K){20}
1: KKKKK (R,K){20}
KKKKKKKKKKKKKKKKKK KKKKA

(K){20}
2: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKKAG

(K){20}
3: KKKKK (K){20}
KKKKKKKKKKKKKKKKKK KKAAG

4: KKK KKKKKKKKKKKKKKKKK KAGG
(K){20}
5: KKK KKKKKKKKKKKKKKKKK AGCGA

AA011843 ck: 86 len: 36 1 Aa011843 Human polypeptide SEQ ID NO 25735.

1

(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKK KKKK

(K){20}
2: K KKKKKKKKKKKKKKKKK KKKK

(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKK

(K){20}
4: KKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
5: KKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
6: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
7: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
8: KKKK KKKKKKKKKKKKKKKKK KKKK

(R){20}
9: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
10: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
11: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
12: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
13: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
14: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
15: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011844 ck: 3574 len: 49 1 Aa011844 Human polypeptide SEQ ID NO 25736.

1

(R,K){20}
(K){20}
16: LLYE KKKKKKKKKKKKKKKKK KKKK

(K){20}
17: LYME KKKKKKKKKKKKKKKKK KKKK

(K){20}
18: YMEK KKKKKKKKKKKKKKKKK KKKK

(K){20}
19: MEKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
20: EKKK KKKKKKKKKKKKKKKKK KKKK

1

(K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
24: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
25: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
26: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
27: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011845 ck: 8375 len: 45 1 Aa011845 Human polypeptide SEQ ID NO 257

(R,K){20}

(K){20}
15: SPSTL KKKKKKKKKKKKKKKKK KKKK

(K){20}
16: PSTLK KKKKKKKKKKKKKKKKK KKKK

(K){20}
17: STLKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
18: TLKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
19: LKKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
20: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
21: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
22: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKKK

(K){20}
24: KKKK KKKKKKKKKKKKKKKKK KKKK

AA011849 ck: 284 len: 36 1 Aa011849 Human polypeptide SEQ ID NO 257

1

(R,K){20}
(K){20}
13: KCLCE KKKKKKKKKKKKKKKKK KKKK

(K){20}
14: CLCEK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}
15: LCEKK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}
16: CEKKK KKKKKKKKKKKKKKKKK KKKK

(R,K){20}
17: EKKKK KKKKKKKKKKKKKKKKK KKKK

1

AA011886 ck: 5282 len: 97 1 Aa011886 Human polypeptide SEQ ID NO 257
(R,K){20}
(K){20}

19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK K
 20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK K
 AAO12105 ck: 6255 len: 80 1 Aaol2105 Human polypeptide SEQ ID NO 25997.
 (R,K){20}
 53: EATSL KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 54: ATSLK KKKKKKKKKKKKKKKKKKKKK KKKKA
 (K){20}
 55: TSLKK KKKKKKKKKKKKKKKKKKKKK KKKAR
 (K){20}
 56: SLKKK KKKKKKKKKKKKKKKKKKKKK KKARG
 (K){20}
 57: LKKKK KKKKKKKKKKKKKKKKKKKKK KARG
 (K){20}
 58: KKKKK KKKKKKKKKKKKKKKKKKKKK ARG
 (K){20}
 AAO12179 ck: 1295 len: 69 1 Aaol2179 Human polypeptide SEQ ID NO 26071.
 (R,K){20}
 35: PFSV KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 36: KESVK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 37: FSVKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 38: SVKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 39: VKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
 (K){20}
 44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKGG
 (K){20}
 45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKGGA
 (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKKK GGGA
 (K){20}
 AAO12180 ck: 9840 len: 67 1 Aaol2180 Human polypeptide SEQ ID NO 26072.
 (R,K){20}

38: CSTLP KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 39: SYLPK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 40: YLPKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 41: LPKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 42: PKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
 (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
 (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKKK K
 (K){20}
 48: KKKKK KKKKKKKKKKKKKKKKKKKKK
 (K){20}
 AAO12187 ck: 4700 len: 60 1 Aaol2187 Human polypeptide SEQ ID NO 260
 (R,K){20}
 30: SCCKL KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 31: CKKLK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 32: KKLKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 33: KLKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 34: LKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
 (K){20}
 37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK
 (K){20}
 38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
 (K){20}
 39: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
 (K){20}
 40: KKKKK KKKKKKKKKKKKKKKKKKKKK K
 (K){20}
 41: KKKKK KKKKKKKKKKKKKKKKKKKKK

1
AA012203 ck: 4093 len: 41 i Aa012203 Human polypeptide SEQ ID NO 26095.
(R,K){20}

6: SMCCL KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

7: WCCLK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

8: CCLKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

9: CLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1
AA012215 ck: 6903 len: 73 i Aa012215 Human polypeptide SEQ ID NO 26107.
(R,K){20}

50: PPPLP KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

51: PDLPK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1
AA012243 ck: 8474 len: 39 i Aa012243 Human polypeptide SEQ ID NO 26135.
(R,K){20}

11: MISFI KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

12: ISFIK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

13: SPFKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

14: FIKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

15: IKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1
AA012250 ck: 1538 len: 69 i Aa012250 Human polypeptide SEQ ID NO 26142.
(R,K){20}

31: SPSNL KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

32: PSNLK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

33: SNLKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

34: NLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

35: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

1
38: KKKKK (K){20}
39: KKKKK (K){20}
40: KKKKK (K){20}
41: KKKKK (K){20}
42: KKKKK (K){20}
43: KKKKK (K){20}
44: KKKKK (K){20}
45: KKKKK (R,K){20}

AA012274 ck: 2739 len: 37 i Aa012274 Human polypeptide SEQ ID NO 261

14: KGLIN KKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

15: GLINK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

16: LINKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

17: INKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

AA012280 ck: 7481 len: 66 i Aa012280 Human polypeptide SEQ ID NO 261

29: HLPPL KKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

30: IPPLK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

31: PPLKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

32: PLKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

33: LKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

34: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

35: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

36: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

37: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

38: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

(K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKK KKKSG
(K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKK KKS GG
(K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKK KSGGG
(K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKK SGGGA
AAO12447 ck: 5017 len: 47 i Aao12447 Human polypeptide SEQ ID NO 26339.
(R,K){20}
(K){20}
14: HLVC E KKKKKKKKKKKKKKKKKKK KRGGA
(K){20}
15: LVCE K KKKKKKKKKKKKKKKKKKK RGAL
(R,K){20}
16: VCE K KKKKKKKKKKKKKKKKKKK GGALK
AAO12476 ck: 837 len: 104 i Aao12476 Human polypeptide SEQ ID NO 26368.
(R,K){20}
(K){20}
65: RFCH Q KKKKKKKKKKKKKKKKKKK KGFEF
(K){20}
66: FCH Q KKKKKKKKKKKKKKKKKKK GFEF
AAO12548 ck: 6973 len: 74 i Aao12548 Human polypeptide SEQ ID NO 26440.
(R,K){20}
(K){20}
38: AVLPL KKKKKKKKKKKKKKKKKKK KKG GG
(K){20}
39: VLPL K KKKKKKKKKKKKKKKKKKK KGGV
(K){20}
40: LPL K KKKKKKKKKKKKKKKKKKK GGGV
AAO12553 ck: 1903 len: 33 i Aao12553 Human polypeptide SEQ ID NO 26445.
(R,K){20}
(K){20}
9: SCCFI KKKKKKKKKKKKKKKKKKK KRGAP
(K){20}
10: CCFIK KKKKKKKKKKKKKKKKKKK RGAP
(R,K){20}
11: CFIK KKKKKKKKKKKKKKKKKKK GAP
AAO13164 ck: 2798 len: 71 i Aao13164 Human polypeptide SEQ ID NO 27056.
(R,K){20}
(R,K){20}
35: RPPLX KKKKKKKKKKKKKKKKKKK KEMFK
(K){20}
36: PPLX KKKKKKKKKKKKKKKKKKK EMFK

1
AAO13576 ck: 4846 len: 99 i Aao13576 Human polypeptide SEQ ID NO 274
(R,K){20}
(K){20}
53: TNNLI KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: NNLIK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
55: NLIK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
56: LIK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
57: IKKK KKKKKKKKKKKKKKKKKKK KKKG
(K){20}
58: KKKK KKKKKKKKKKKKKKKKKKK KKG G
(K){20}
59: KKKK KKKKKKKKKKKKKKKKKKK KGGP
(K){20}
60: KKKK KKKKKKKKKKKKKKKKKKK GGGP
AAO13785 ck: 6241 len: 100 i Aao13785 Human polypeptide SEQ ID NO 276
(R,K){20}
(K){20}
41: LYAP KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
42: YAP K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
43: AP K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
44: PPK K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
45: PK K KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
46: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
47: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
51: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
52: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
53: KKK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
54: KKK KKKKKKKKKKKKKKKKKKK KKKK


```
55: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKAAG
56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKAAG
57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KAGGG

AAU18049 ck: 7016 len: 315 1 Aau18049 Human immunoglobulin polypeptide S
(R,K){20}
(R,K){20}
273: QVFAP KKKKKKKKKKKKKKKKKKK KAGRS
(K){20}
274: VFAPR KKKKKKKKKKKKKKKKKKK KGRSR

AAU18049 ck: 7611 len: 315 1 Aau18049 Human immunoglobulin polypeptide S
(R,K){20}
(R,K){20}
273: QVFAP KKKKKKKKKKKKKKKKKKK KAGRS
(K){20}
274: VFAPR KKKKKKKKKKKKKKKKKKK KGRSR

AAM14119 ck: 5383 len: 86 1 Aam14119 Peptide #553 encoded by probe for
(R,K){20}
(R,K){20}
15: RRRRG RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
16: RRRGR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
17: RRGRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
18: RGRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
19: GRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
20: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
21: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
22: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
23: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
24: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
25: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
26: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
27: RRRRR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
```

```
28: KKKKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
29: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
30: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
31: RRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
32: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
33: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
34: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
35: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
36: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
37: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
38: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
39: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
40: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
41: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
42: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
43: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
44: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
45: KRRKK KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}

AAM14961 ck: 1334 len: 86 1 Aam14961 Peptide #1395 encoded by probe
(R,K){20}
(R,K){20}
57: EEEEG RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
58: EEEGR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
59: EEEGR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
60: EEEGR RRRRRKKKKKKKKKKKKKK KRRR
(R,K){20}
61: GRRRR KKKKKKKKKKKKKKKKKKK KRRR
(R,K){20}
```

62: RRRKK KRRKKKKKKKKKKKKKK KKKKK
(R,K){20}
63: RRRKK KRRKKKKKKKKKKKKKK KKKK
(R,K){20}
64: RRRKK KRRKKKKKKKKKKKKKK KKK
(R,K){20}
65: KRRKK KRRKKKKKKKKKKKKKK KK
(R,K){20}
66: KRRKK KRRKKKKKKKKKKKKKK K
(K){20}
67: KRRKK KRRKKKKKKKKKKKKKK

AAm15038 ck: 9082 len: 167 1 Aam15038 Peptide #1472 encoded by probe for

33: EEEGG (R,K){20}
(R,K){20}
34: EGGGR (R,K){20}
(R,K){20}
35: GGGRR (R,K){20}
(R,K){20}

AAm15527 ck: 2276 len: 89 1 Aam15527 Peptide #1961 encoded by probe for

23: EEEEE (R,K){20}
(R,K){20}
46: KEEEE (K){20}
(K){20}
47: KEEEE (K){20}
(K){20}
48: EEEEE (R,K){20}
(R,K){20}
49: EEEEE (R,K){20}
(R,K){20}

AAm15826 ck: 1939 len: 130 1 Aam15826 Peptide #2260 encoded by probe for

42: EGGRE (R,K){20}
(R,K){20}
43: GKKER (R,K){20}
(R,K){20}
44: RKEER (R,K){20}
(R,K){20}
45: KERRR (R,K){20}
(R,K){20}
46: ERRRR (R,K){20}
(R,K){20}
47: RRRRR (R,K){20}
(R,K){20}
48: RRRRR (R,K){20}
(R,K){20}

49: RRRRR (R,K){20}
(R,K){20}
50: RRRRR (R,K){20}
(R,K){20}
51: RRRRR (R,K){20}
(R,K){20}
52: RRRRR (R,K){20}
(R,K){20}
53: RRRRR (R,K){20}
(R,K){20}
54: RRRRR (R,K){20}
(R,K){20}

AAm16123 ck: 3607 len: 88 1 Aam16123 Peptide #2557 encoded by probe

39: EKKRE (R,K){20}
(K){20}
40: RKRER (K){20}
(K){20}
41: KREER (K){20}
(K){20}
42: REKKR (K){20}
(K){20}
43: EKKKK (K){20}
(K){20}
44: KKKKK (K){20}
(K){20}
45: KKKKK (K){20}
(K){20}
46: KKKKK (K){20}
(K){20}
47: KKKKK (K){20}
(K){20}
48: KKKKK (K){20}
(K){20}
49: KKKKK (K){20}
(K){20}
50: KKKKK (K){20}
(K){20}
51: KKKKK (K){20}
(K){20}
52: KKKKK (K){20}
(K){20}
53: KKKKK (K){20}
(K){20}
54: KKKKK (K){20}
(K){20}
55: KKKKK (K){20}
(K){20}

1

```

56: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
57: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
58: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
59: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
60: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
61: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
62: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
63: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
64: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
65: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
66: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
67: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
68: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
69: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

AAM16691 ck: 3937 len: 85 1 Aam16691 Peptide #3125 encoded by probe for
(R,K){20}
(K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK KKKKK
2: K (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
3: KK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
4: KKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
5: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
6: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
7: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
8: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK
9: KKKKK (K){20} KKKKKKKKKKKKKKKKKKKKK KKKKK

```

1

```

10: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
11: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
12: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
13: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
14: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
15: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
16: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
17: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
18: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
19: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
20: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
21: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
22: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
23: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
24: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
25: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
26: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
27: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
28: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
29: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
30: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
31: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
32: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
33: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
34: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
35: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
36: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
37: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
38: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
39: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK
40: KKKKK (K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

1

```
41: KKKKK (R,K){20} KKKKK
42: KKKKK (R,K){20} KKKKK
43: KKKKK (R,K){20} KKKKK
44: KKKKK (R,K){20} KKKKK
45: KKKKK (R,K){20} KKKKK
46: KKKKK (R,K){20} KKKKK
47: KKKKK (K){20} KKKKK
48: KKKKK (K){20} KKKKK
49: KKKKK (K){20} KKKKK
AAM18532 ck: 1560 len: 88 1 Aam18532 Peptide #4966 encoded by probe for
43: RRRGG (R,K){20} RRRGG
44: RRRGG (R){20} RRRGG
45: RRRGG (R){20} RRRGG
46: RRRGG (R){20} RRRGG
47: GRRRR (R){20} GRRRR
48: RRRRR (R){20} RRRRR
49: RRRRR (R){20} RRRRR
50: RRRRR (R){20} RRRRR
51: RRRRR (R){20} RRRRR
52: RRRRR (R){20} RRRRR
53: RRRRR (R){20} RRRRR
54: RRRRR (R){20} RRRRR
55: RRRRR (R){20} RRRRR
56: RRRRR (R){20} RRRRR
```

1

```
57: RRRRR (R){20} RRRRR
58: RRRRR (R){20} RRRRR
59: RRRRR (R){20} RRRRR
60: RRRRR (R){20} RRRRR
61: RRRRR (R){20} RRRRR
62: RRRRR (R){20} RRRRR
AAM20089 ck: 8343 len: 66 1 Aam20089 Peptide #6523 encoded by probe
6: ETERE (R,K){20} ETERE
7: TERER (R,K){20} TERER
8: EREKK (R,K){20} EREKK
9: REKKK (R,K){20} REKKK
AAM21941 ck: 4895 len: 51 1 Aam21941 Peptide #8375 encoded by probe
18: LRRPM (R,K){20} LRRPM
19: FKPMR (R,K){20} FKPMR
20: KPMRK (R,K){20} KPMRK
AAM26526 ck: 5383 len: 86 1 Aam26526 Peptide #563 encoded by probe f
15: RRRGG (R,K){20} RRRGG
16: RRRGG (R,K){20} RRRGG
17: RRRGG (R,K){20} RRRGG
18: RRRGG (R,K){20} RRRGG
19: GRRRR (R,K){20} GRRRR
20: RRRRR (R,K){20} RRRRR
21: RRRRR (R,K){20} RRRRR
```

22: RRRKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
23: RRRKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
24: RKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
25: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
26: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
27: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
28: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
29: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
30: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
31: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
32: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
33: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
34: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
35: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
36: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
37: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
38: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
39: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
40: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
41: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
42: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
43: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
44: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR
45: KKKKK (R,K){20} KKKKKKKKKKKKKRRRRR RRRRR

1

AAM27395 ck: 1334 len: 86 ! Aam27395 Peptide #1432 encoded by probe
57: EEEEG (R,K){20} RRRRRKKKKKKKKKKKKKK KKKKK
58: EEEGR (R,K){20} RRRRRKKKKKKKKKKKKKK KKKKK
59: EEEGR (R,K){20} RRRRRKKKKKKKKKKKKKK KKKKK
60: EEEGR (R,K){20} RRRRRKKKKKKKKKKKKKK KKKKK
61: GRRRK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
62: RRRKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
63: RRRKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
64: RRRKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
65: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
66: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK
67: KKKKK (R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

1

AAM27479 ck: 9082 len: 167 ! Aam27479 Peptide #1516 encoded by probe
33: EEEGR (R,K){20} RRRRRKKKKKKKKKKKKKK RRRRR
34: EEEGR (R,K){20} RRRRRKKKKKKKKKKKKKK RRRRR
35: GGGRR (R,K){20} RRRRRKKKKKKKKKKKKKK RRRRR

1

AAM28016 ck: 2276 len: 89 ! Aam28016 Peptide #2053 encoded by probe
23: EEEEE (R,K){20} KKKKKKKKKKKKKKKKKKK EEEEE
46: KKEEE (R,K){20} KKKKKKKKKKKKKKKKKKK KKEEE
47: KKEEK (R,K){20} KKKKKKKKKKKKKKKKKKK KKEEK
48: EEEKK (R,K){20} KKKKKKKKKKKKKKKKKKK EEEKK
49: EEEKK (R,K){20} KKKKKKKKKKKKKKKKKKK EEEKK

1

AAM28336 ck: 1939 len: 130 ! Aam28336 Peptide #2373 encoded by probe
(R,K){20} KKKKKKKKKKKKKKKKKKK KKKKK

```

42: EGRKE RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
43: GKRR RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
44: KERR RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
45: KERR RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
46: ERRR RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
47: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
48: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
49: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
50: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
51: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
52: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
53: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}
54: RRRR RRRRRRRRRRRRRRRR RRRR
      (R,K){20}

AAM28616 cK: 3607 len: 88 1 Aam28616 Peptide #2653 encoded by probe for
      (R,K){20}
39: EKRRE RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
40: KRREK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
41: KRREK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
42: REKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
43: EKRRK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
44: KKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
45: KKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
46: KKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
47: KKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}
48: KKKK RRRRRRRRRRRRRRRRRR RRRR
      (R,K){20}

```

```

49: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
50: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
51: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
52: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
53: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
54: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
55: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
56: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
57: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
58: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
59: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
60: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
61: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
62: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
63: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
64: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
65: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
66: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
67: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
68: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
69: KKKK KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}

AAM29178 ck: 3937 len: 85 1 Aam29178 peptide #3215 encoded by probe
      (R,K){20}
      (K){20}
1: KKKKKKKKKKKKKKKKKKK KKKK
      (K){20}
2: K KKKKKKKKKKKKKKKKK KKKK
      (K){20}

```

3: KK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
EEXX
AAM30989 ck: 1560 len: 88 1 Aam30989 Peptide #5026 encoded by probe for
(R,K){20}
43: EERRG RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
44: EPRGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
45: PRGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
46: RGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
47: GRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
48: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
49: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
50: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
51: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
52: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
53: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
54: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
55: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
56: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

57: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
58: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
59: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
60: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
61: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
62: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}
NTNE
AAM31173 ck: 2324 len: 36 1 Aam31173 Peptide #5210 encoded by probe
(R,K){20}
16: KERKT KKKRRRRRRRRRRRRRRRRRRRRRRRR
(R,K){20}
17: ERKT KKKRRRRRRRRRRRRRRRRRRRRRRRR
(R,K){20}

AAM33953 ck: 8343 len: 66 1 Aam33953 Peptide #7990 encoded by probe
(R,K){20}

6: ETERE KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

7: TEREX KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

8: EREKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

9: REKKK KKKKKKKKKKKKKKKKKKKKKKKKKKK
(R,K){20}

AAM36456 ck: 2394 len: 57 1 Aam36456 Peptide #10493 encoded by probe
(R,K){20}

20: EEEEG RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

21: EEGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

22: EGR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

23: EGRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

24: GRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

25: RRRR RRRRRRRRRRRRRRRRRRRRRRRRRRR
(R){20}

AAM37018 ck: 4228 len: 24 1 Aam37018 Peptide #11055 encoded by probe
(R,K){20}

1: RRRRRKKKKRRRRRRRRRRRRRRRRRRRRRRRR
(R,K){20}

1

```
2: R RRRRRRRRRRRRRRRRRRRR KRT
   (R,K){20}
3: RR RRRRRRRRRRRRRRRRRRRR RT
   (R,K){20}
4: RRR RRRRRRRRRRRRRRRRRRRR T
   (R,K){20}

AAM37794 ck: 3301 len: 52 1 Aam37794 Peptide #11831 encoded by probe fd
   (R,K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
```

1

1

```
AAM38273 ck: 4895 len: 51 1 Aam38273 Peptide #12310 encoded by probe
   (R,K){20}
18: LKPKK KKKKKKKKKKKKKKKKKRR KKLTT
   (R,K){20}
19: FKPKR KKKKKKKKKKKKKKKKKRR KKLTT
   (R,K){20}
20: KPMKK RRRKKKKKKKKKKKKRRKK LTTT

AAU04283 ck: 8137 len: 45 1 Aau04283 Trimeric fusogenic peptide #2 u
   (R,K){20}
4: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
5: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
6: YKA KKKKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
   (R,K){20}
```



```
(K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKK
   (K){20}
24: KKKK KKKKKKKKKKKKKKKKK WK
   (K){20}
AAU04285 ck: 4361 len: 59 1 Au04285 Nuclear ligand #2 used in nucleic
   (R,K){20}
18: ADYKA KKKKKKKKKKKKKKKKK KKKK
   (K){20}
19: PYAK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
20: YRAK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
21: KAKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
22: AKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
24: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
25: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
26: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
27: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
28: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
29: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
30: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
31: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
32: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
33: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
34: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
35: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
36: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
37: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
38: KKKK KKKKKKKKKKKKKKKKK WK
```

```
1
AAU04287 ck: 4925 len: 100 1 Au04287 Poly-L-lysine used in nucleic
   (R,K){20}
1: KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
2: K KKKKKKKKKKKKKKKKK KKKK
   (K){20}
3: K KKKKKKKKKKKKKKKKK KKKK
   (K){20}
4: K KKKKKKKKKKKKKKKKK KKKK
   (K){20}
5: KKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
6: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
9: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
10: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
11: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
12: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
13: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
14: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
15: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
16: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
17: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
18: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
19: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
20: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
21: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
22: KKKK KKKKKKKKKKKKKKKKK KKKK
   (K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKKK
```

24: KKKK (K){20}
 25: KKKK (K){20}
 26: KKKK (K){20}
 27: KKKK (K){20}
 28: KKKK (K){20}
 29: KKKK (K){20}
 30: KKKK (K){20}
 31: KKKK (K){20}
 32: KKKK (K){20}
 33: KKKK (K){20}
 34: KKKK (K){20}
 35: KKKK (K){20}
 36: KKKK (K){20}
 37: KKKK (K){20}
 38: KKKK (K){20}
 39: KKKK (K){20}
 40: KKKK (K){20}
 41: KKKK (K){20}
 42: KKKK (K){20}
 43: KKKK (K){20}
 44: KKKK (K){20}
 45: KKKK (K){20}
 46: KKKK (K){20}
 47: KKKK (K){20}

48: KKKK (K){20}
 49: KKKK (K){20}
 50: KKKK (K){20}
 51: KKKK (K){20}
 52: KKKK (K){20}
 53: KKKK (K){20}
 54: KKKK (K){20}
 55: KKKK (K){20}
 56: KKKK (K){20}
 57: KKKK (K){20}
 58: KKKK (K){20}
 59: KKKK (K){20}
 60: KKKK (K){20}
 61: KKKK (K){20}
 62: KKKK (K){20}
 63: KKKK (K){20}
 64: KKKK (K){20}
 65: KKKK (K){20}
 66: KKKK (K){20}
 67: KKKK (K){20}
 68: KKKK (K){20}
 69: KKKK (K){20}
 70: KKKK (K){20}
 71: KKKK (K){20}
 72: KKKK (K){20}

1

```
73: KKKKK (K){20} KKKKK
74: KKKKK (K){20} KKKKK
75: KKKKK (K){20} KKKKK
76: KKKKK (K){20} KKKKK
77: KKKKK (K){20} KKKKK
78: KKKKK (K){20} KKKKK
79: KKKKK (K){20} KKKKK
80: KKKKK (K){20} K
81: KKKKK (K){20} KKKKK
AAM01857 ck: 5383 len: 86 1 Aam01857 Peptide #539 encoded by probe for
15: RRRRG (R,K){20} RRRRG
16: RRRGR (R,K){20} RRRGR
17: RRRGR (R,K){20} RRRGR
18: RRRGR (R,K){20} RRRGR
19: GRRRR (R,K){20} RRRRR
20: RRRRR (R,K){20} RRRRR
21: RRRRK (R,K){20} RRRRR
22: RRRKK (R,K){20} RRRRR
23: RRRKK (R,K){20} RRRRR
24: RRRKK (R,K){20} RRRRR
25: KRRKK (R,K){20} RRRRR
26: KRRKK (R,K){20} RRRRR
27: KRRKK (R,K){20} RRRRR
28: KRRKK (R,K){20} RRRRR
```

1

```
29: KKKKK (R,K){20} RRRRR
30: KRRKK (R,K){20} RRRRR
31: KRRKK (R,K){20} RRRRR
32: KRRKK (R,K){20} RRRRR
33: KRRKK (R,K){20} RRRRR
34: KRRKK (R,K){20} RRRRR
35: KRRKK (R,K){20} RRRRR
36: KRRKK (R,K){20} RRRRR
37: KRRKK (R,K){20} RRRRR
38: KRRKK (R,K){20} RRRRR
39: KRRKK (R,K){20} RRRRR
40: KRRKK (R,K){20} RRRRR
41: KRRKK (R,K){20} RRRRR
42: KRRKK (R,K){20} RRRRR
43: RRRRR (R,K){20} RRRRR
44: RRRRR (R,K){20} RRRRR
45: RRRRR (R,K){20} RRRRR
AAM02687 ck: 1334 len: 86 1 Aam02687 Peptide #1369 encoded by probe
57: EEEEG (R,K){20} KKKKK
58: EEEGR (R,K){20} KKKKK
59: EEEGR (R,K){20} KKKKK
60: EEEGR (R,K){20} KKKKK
61: GRRRK (R,K){20} KKKKK
62: RRRKK (R,K){20} KKKKK
```

```

(R,K){20}
63: RRRR RRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
64: RRRR RRRRRRRRRRRRRRRRRR KKK
(R,K){20}
65: KKKR KRRRRRRRRRRRRRRRRR KK
(R,K){20}
66: KKKR RRRRRRRRRRRRRRRRRR K
(K){20}
67: KRRR RRRRRRRRRRRRRRRRRR

AAM02768 ck: 9082 len: 167 ! Aam02768 Peptide #1450 encoded by probe for
(R,K){20}
33: EGRG RRRRRRRRRRRRRRRRRR RRGG
(R,K){20}
34: EGRG RRRRRRRRRRRRRRRRRR RGGG
(R,K){20}
35: GRGR RRRRRRRRRRRRRRRRRR GGGG

AAM03278 ck: 2276 len: 89 ! Aam03278 Peptide #1960 encoded by probe for
(R,K){20}
23: EEEE RRRRRRRRRRRRRRRRRR EEEK
(K){20}
46: KKEE RRRRRRRRRRRRRRRRRR KRKE
(K){20}
47: KKEE RRRRRRRRRRRRRRRRRR RKEE
(R,K){20}
48: EEEK RRRRRRRRRRRRRRRRRR KEEE
(R,K){20}
49: EEEK RRRRRRRRRRRRRRRRRR EEEE

AAM03564 ck: 1939 len: 130 ! Aam03564 Peptide #2246 encoded by probe for
(R,K){20}
42: EGKE RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
43: GKER RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
44: KEER RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
45: KEER RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
46: EERR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
47: RRRR RRRRRRRRRRRRRRRRRR RRRR
(R,K){20}
48: RRRR RRRRRRRRRRRRRRRRRR RRRR

```

```

(R,K){20}
49: RRRR RRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
50: RRRR RRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
51: RRRR RRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
52: RRRR RRRRRRRRRRRRRRRRRR KKKK
(R,K){20}
53: RRRR RRRRRRRRRRRRRRRRRR KEEE
(R,K){20}
54: RRRR RRRRRRRRRRRRRRRRRR EEEE

AAM03848 ck: 3607 len: 88 ! Aam03848 Peptide #2530 encoded by probe
(R,K){20}
39: ERKE RRRRRRRRRRRRRRRRRR KKKK
(K){20}
40: RREK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
41: KREK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
42: REKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
43: EKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
44: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
45: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
46: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
47: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
48: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
49: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
50: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
51: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
52: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
53: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
54: KKKK RRRRRRRRRRRRRRRRRR KKKK
(K){20}
55: KKKK RRRRRRRRRRRRRRRRRR KKKK

```

(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKK

AAm04408 ck: 3937 len: 85 1 Aam04408 Peptide #3090 encoded by probe for

(R,K){20}
1: KKKKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
2: K KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
4: KKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
5: KKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
6: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
7: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
8: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK
(K){20}

(K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKEEE
(K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK EEEEE

AAm06100 ck: 1560 len: 88 1 Aam06100 Peptide #4782 encoded by probe

(R,K){20}
43: RRRRG RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
44: RRRGR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
45: RRGRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
46: RGRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
47: GRGRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
54: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
58: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
59: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR
(R){20}

AAg73687 ck: 3063 len: 25 1 Aag73687 Human colon cancer antigen prot

```
1
(R,K){20}
8: MDDTF KKKKKKKKKKKKKKKKKKK KK
(K){20}
9: MMTFK KKKKKKKKKKKKKKKKKKK X
(K){20}

AAG73729 ck: 783 len: 83 1 Aag73729 Human colon cancer antigen protein
(R,K){20}
49: LCPCE KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: GPCEK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
51: PCEKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
52: CEKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: EKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAG73810 ck: 3374 len: 88 1 Aag73810 Human colon cancer antigen protein
(R,K){20}
44: FGDTX KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
45: GQTXK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
46: QTXKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
47: TXKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
48: XKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

1
(R,K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAG73895 ck: 1887 len: 43 1 Aag73895 Human colon cancer antigen prot
(R,K){20}
18: VRPRV KKKKKKKKKKKKKKKKKKK KKKKK
(R,K){20}
19: RPRVR KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
20: PRVRK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
21: RVRKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
22: VRKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}

AAG74218 ck: 8659 len: 104 1 Aag74218 Human colon cancer antigen prot
(R,K){20}
75: PLGGQ KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
76: LGGQK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
77: GGQKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
78: GQKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
79: QKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
```

AAg74527 ck: 2664 len: 40 : Aag74527 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

9: CLLLY KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

10: LLLYK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

11: LLYKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

12: LYKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

13: YKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

14: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

15: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74650 ck: 1596 len: 69 : Aag74650 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

36: LQCRQ KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

37: QCRQK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

38: CRQKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

39: RQKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

40: QKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

41: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

42: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

43: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

44: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

45: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

46: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74793 ck: 8497 len: 152 : Aag74793 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

122: SSHQK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

123: SHQKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg74907 ck: 1215 len: 98 : Aag74907 Human colon cancer antigen protein

1 (R,K){20}

(K){20}

57: NLRRK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

58: LRKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

59: RKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

60: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

61: EKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

62: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

63: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

64: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

65: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

66: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

67: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

68: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

69: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

70: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

71: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

72: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

73: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

74: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

75: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

(K){20}

76: KKKKK KKKKKKKKKKKKKKKKKKKKKKKKK

AAg75215 ck: 3913 len: 155 : Aag75215 Human colon cancer antigen protein

1 (R,K){20}

```

135: RSGAP      (K) {20}
          KKKKKKKKKKKKKKKKKKKKK K
136: SSAPK      (K) {20}
          KKKKKKKKKKKKKKKKKKKKK

```

1
AAAG75986 CR: 4235 len: 71 | Aag75986 Human colon cancer antigen protein
(R,K){20}
(K){20}
47: KKKXX KKKKKKKKKKKKKKKKKKK KXGXX
(K){20}
48: KKKXX KKKKKKKKKKKKKKKKKKK KXGXX

```

1
AAE01796  ck: 4416  len: 72  | Aae01796 Human gene 27 encoded secreted pro
(R,K){20}
47: LEPL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: PTELK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: TELK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
50: ELKK KKKKKKKKKKKKKKKKKKK KKK
(K){20}
51: LKKK KKKKKKKKKKKKKKKKKKK KK
(K){20}
52: KKKK KKKKKKKKKKKKKKKKKKK K
(K){20}
53: KKKK KKKKKKKKKKKKKKKKKKK
(K){20}

```

```

1
AAE01648 ck: 5584 len: 73 i: Aae01648 Human gene 27 encoded secreted pro
(R,K){20}
47: LPTEL KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
48: PTELK KKKKKKKKKKKKKKKKKKK KKKK
(K){20}
49: TELKK KKKKKKKKKKKKKKKKKKK KKKKI
(K){20}
50: ELKKK KKKKKKKKKKKKKKKKKKK KKKI
(K){20}
51: LKKKK KKKKKKKKKKKKKKKKKKK KKI
(K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKK KI
(K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKK I
(K){20}

```

AAB90574 ck: 1431 len: 530 ; Aab90574.Human secreted protein, SEQ ID NO:
1
(R,K){20}
(K){20}
511: LHAAP KKKKKKKKKKKKKKKKK

AA045846 ck: 8137 len: 45 1 Aab45846 Nucleic acid transporter system

```
(R,K){20}
4: YKA KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
5: YKAK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
6: YKAKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
7: KAKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
8: AKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
9: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
   (K){20}
```

1

AAB45848	ck:	4361	len:	59	!	Aab45848	Nucleic acid transporter system
(R,K)[20]							
(K)[20]							
18: APYKA	KKKKKKKKKKKKKKKKKKKK	KKKK					

19: PYAK (K){20} KKKK
(K){20}
20: YKAK KKKKKKKKKKKKKKKKK KKKK
(K){20}
21: KAKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
22: AKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
23: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
25: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
26: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
27: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
28: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
29: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
30: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
31: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
32: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
33: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
34: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
35: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
36: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
37: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}
38: KKKK KKKKKKKKKKKKKKKKK KKKK
(K){20}

AAb45850 ck: 4925 len: 100 1 Aab45850 Nucleic acid transporter system pe

1 (R,K){20} KKKKKKKKKKKKKKKKK KKKK
(K){20}
2: K (K){20} KKKKKKKKKKKKKKKKK KKKK
(K){20}
3: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}

4: KK (K){20} KKKKKKKKKKKKKKKKK KKKK
(K){20}
5: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
6: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
7: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
8: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
9: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
10: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
11: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
12: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
13: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
14: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
15: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
16: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
17: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
18: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
19: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
20: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
21: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
22: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
23: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
24: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
25: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
26: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}
27: KK KKKKKKKKKKKKKKKKK KKKK
(K){20}

28: kkkk (K){20} kkkk
29: kkkk (K){20} kkkk
30: kkkk (K){20} kkkk
31: kkkk (K){20} kkkk
32: kkkk (K){20} kkkk
33: kkkk (K){20} kkkk
34: kkkk (K){20} kkkk
35: kkkk (K){20} kkkk
36: kkkk (K){20} kkkk
37: kkkk (K){20} kkkk
38: kkkk (K){20} kkkk
39: kkkk (K){20} kkkk
40: kkkk (K){20} kkkk
41: kkkk (K){20} kkkk
42: kkkk (K){20} kkkk
43: kkkk (K){20} kkkk
44: kkkk (K){20} kkkk
45: kkkk (K){20} kkkk
46: kkkk (K){20} kkkk
47: kkkk (K){20} kkkk
48: kkkk (K){20} kkkk
49: kkkk (K){20} kkkk
50: kkkk (K){20} kkkk
51: kkkk (K){20} kkkk

52: kkkk (K){20} kkkk
53: kkkk (K){20} kkkk
54: kkkk (K){20} kkkk
55: kkkk (K){20} kkkk
56: kkkk (K){20} kkkk
57: kkkk (K){20} kkkk
58: kkkk (K){20} kkkk
59: kkkk (K){20} kkkk
60: kkkk (K){20} kkkk
61: kkkk (K){20} kkkk
62: kkkk (K){20} kkkk
63: kkkk (K){20} kkkk
64: kkkk (K){20} kkkk
65: kkkk (K){20} kkkk
66: kkkk (K){20} kkkk
67: kkkk (K){20} kkkk
68: kkkk (K){20} kkkk
69: kkkk (K){20} kkkk
70: kkkk (K){20} kkkk
71: kkkk (K){20} kkkk
72: kkkk (K){20} kkkk
73: kkkk (K){20} kkkk
74: kkkk (K){20} kkkk
75: kkkk (K){20} kkkk
76: kkkk (K){20} kkkk

```

77: KKKKK KKKKKKKKKKKKKKKKKKK KKKK
      (<f>)[20]

78: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK
      (<f>)[20]

79: KKKKK KKKKKKKKKKKKKKKKKKKKK KK
      (<f>)[20]

80: KKKKK KKKKKKKKKKKKKKKKKKKKK K
      (<f>)[20]

81: KKKKK KKKKKKKKKKKKKKKKKKKKK
      (<f>)[20]

```

1
(R,K) [20]

114:	TOLRQ	KKKKKKKKKKKKKKKKKKKK	KKKK
	(R,K){20}	(K){20}	
115:	OLRQK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
116:	LROK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
117:	RQKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
118:	OQKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
119:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
120:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
121:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
122:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
123:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
124:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
125:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
126:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
127:	KKKK	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		
128:	KKKK	KKKKKKKKKKKKKKKKKKKK	XXXX
	(K){20}		
ABD4d30	OK: 5275	len: 38	! Abb
	(R,K){20}		
4:	MAQ	KKKKKKKKKKKKKKKKKKKK	KKKK
	(K){20}		

1

25: KKKKK KKKKKKKKKKKKKKKKKKKKK GG
AAU69736 ck: 5862 len: 50 ! Aau69736 Cell death protective sequence CNI

(R,K){20}
(R,K){20}

11: ESALG KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

12: SALGR KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

13: ALGRK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

14: LGRKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

15: GRKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

16: RKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

20: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

21: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

22: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

23: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

24: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

25: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

26: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

27: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}

29: KKKKK KKKKKKKKKKKKKKKKKKKKK GR

Databases searched:
EMBL, Release 7.0, Released on 28Apr2002, Formatted on 2May2002

Total finds: 7,093
Total length: 114,001,827
Total sequences: 766,495
CPU time: 12:26.84